2019-11-04

Compact Cas9s and Their Natural Inhibitors for Genome Editing

Alireza Edraki
University of Massachusetts Medical School

Compact Cas9s and Their Natural Inhibitors for Genome Editing

A Dissertation Presented

By

Alireza Edraki

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

11/04/2019

RNA Therapeutics Institute – Interdisciplinary Graduate Program

Compact Cas9s and Their Natural Inhibitors for Genome Editing

A Dissertation Presented

By

Alireza Edraki

This work was undertaken in the Graduate School of Biomedical Sciences

RNA Therapeutics Institute – Interdisciplinary Graduate Program

Under the mentorship of

Erik J. Sontheimer, Ph.D., Thesis Advisor

Thomas Fazzio, Ph.D., Member of Committee

Christopher Sassetti, Ph.D., Member of Committee

Wen Xue, Ph.D., Member of Committee

Samuel H. Sternberg Ph.D., External Member of Committee

Victor Ambros, Ph.D., Chair of Committee

Mary Ellen Lane, Ph.D., Dean of the Graduate School of Biomedical Sciences

11/04/2019

Acknowledgements

There are many people who have significantly contributed to my scientific and personal career throughout graduate school. First and foremost, I would like to thank Erik Sontheimer for his mentorship, guidance and friendship. From day one, Erik's support helped me gain confidence in myself and my science, and I am grateful to him for providing the ideal environment for my growth as a researcher. Erik is a fantastic scientist, and every time I walked out of his office, I felt elated and motivated. He gave me the freedom to indulge my curiosity, to experiment with my ideas and to make mistakes, and I am grateful for that. His utmost dedication to my career has been exceptional, he provided ample support for me to prepare for my next steps in life, and I feel extremely lucky to have been mentored by Erik.

I would like to extend my acknowledgments to the members of my thesis committee: Victor Ambros, Thomas Fazzio, Christopher Sassetti and Samuel Sternberg. Beyond his role as my committee chair, Victor has always been willing and available to talk and offer advice with many decisions throughout my PhD career. Victor was one of the reasons I joined UMMS and continues to be my science hero and role model. In addition, I want to thank Craig Mello, Scot Wolfe, Darryl Conte and Emily Mohn for their guidance and support in making career and scientific decisions.

I would also like to thank all current and past members of the Sontheimer lab. They created an environment that made science fun, and the long hours in the lab were filled with laughs. I want to thank Nadia for her mentorship, teaching me the very basics of molecular biology. I want to thank Daniel and Raed for the countless hours of

discussion, some of which extended to the realm of automotive mechanics; and thanks to Jooyoung for being a great friend and collaborator. A special thanks to my friend and mentor Aamir, who continues to be a supportive friend and a role model, and I would not have been able to learn as much during my PhD without him. I want to thank members of the Xue and Rivera labs here at UMass for our collaborations. I also would like to extend my gratitude to Ildar of the Zamore lab and Andrew of the Moore lab who spent countless hours to teach me the basics. Thanks to Greg and Dan from the Mello lab for numerous scientific and pseudoscientific conversations. Outside of UMass, I want to thank the numerous collaborators that have enriched my research and scientific development. In particular, I want to thank April Pawluk, who has been a tremendous friend and colleague.

Last but not least, I want to thank my family for their endless support. I would have never made it through my PhD without Sami, and I am thankful for her intellectual and emotional support. I want to thank Mahdi for being a supportive brother, cycling buddy and tennis opponent. I would like to extend my gratitude to my mom and dad for their endless support. Even though my parents are more than 8000 miles away, their encouragement and advice made my graduate career possible, and I am forever indebted to them.

I am grateful the National Institute of Neurological Disorders and Stroke (NINDS) for their support.

ABSTRACT

Recent advances with the bacterial CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) defense system as genome editing tools have opened a new avenue for targeting disease-causing mutations. The programmability of the Cas9 endonuclease by RNA makes it a potentially powerful therapeutic tool to correct such mutations. The CRISPR-Cas9 system consists of a Cas9 endonuclease that is guided by RNA (sgRNA) to create double-stranded breaks in a target DNA segment complementary to the guide. This process is dependent on a 2-8 nucleotide sequence (called PAM) that is adjacent to the target and functions as a Cas9 binding signal. Each Cas9 ortholog recognizes a unique PAM.

However, factors such as the size of Cas9 or the frequency of its PAM sequence in the genome have hindered its clinical use. The Cas9 from *Streptococcus pyogenes* (SpyCas9) is commonly used in research because its PAM (NGG, where "N" symbolizes any nucleotide) is present every ~8 bp in the genome, providing robust targeting potential. However, it is too large to fit into typical viral vectors used for *in vivo* delivery, namely adeno-associated vectors (AAV). While several Cas9 orthologs have been characterized, none satisfied the need for a compact, accurate Cas9 with a short PAM.

In this thesis, we use two approaches to identify new compact Cas9 orthologs with small PAMs, one using anti-CRISPR proteins and one by searching through closely related Cas9s. First, we use the presence of anti-CRISPRs (naturally occurring,

phage-encoded peptides that inhibit CRISPR-Cas9 described in chapter 2) in a genome as indicators of Cas9s that may be highly active. These orthologs come with the added advantage of having inhibitors that can be used as off-switches. We characterize four Cas9s that are targeted by anti-CRISPR proteins and show that they recognize diverse PAMs *in vitro*. One of the four Cas9's, namely HpaCas9 from *Haemophilus parainfluenzae*, induces efficient genome editing in mammalian cells. However, its long N_4GATTT PAM does not satisfy the short PAM criterion.

For our second approach, we asked whether closely related Cas9 orthologs with drastically different PAM-interacting domains (PIDs, the domain responsible for PAM recognition) recognize different PAMs, and if so, can be used for genome editing. To this end, we exploited natural variation in the PID of closely related Cas9s to identify a compact ortholog from *Neisseria meningitidis* (Nme2Cas9). Nme2Cas9 recognizes a simple dinucleotide PAM (N_4CC) that provides a high target site density. All-in-one AAV delivery of Nme2Cas9 with a guide RNA into adult mouse liver produces efficient genome editing and reduced serum cholesterol with exceptionally high specificity. We further expand our single-AAV platform to pre-implanted zygotes for streamlined generation of genome-edited mice. Finally, we show preliminary data on how CRISPR-Cas9 can be used for therapeutic genome editing for Amytrophic Lateral Sclerosis. Our new findings promise to accelerate the development of genome editing tools for biomedical and therapeutic applications.

TABLE OF CONTENTS

Chapter 1: Introduction

Chapter 2: Characterization of Phage-encoded inhibitors of CRISPR-Cas9

Chapter 3: Characterization of compact Cas9 orthologs targeted by Acrs

LIST OF FIGURES

Chapter 1: Introduction

Chapter 2: Phage-encoded inhibitors of CRISPR-Cas9

Chapter 3: Characterization of compact Cas9 orthologs

Chapter 4: Characterization of compact Cas9 orthologs

Chapter 5: Therapeutic Genome Editing for Amyotrophic Lateral Sclerosis

LIST OF TABLES

Chapter 1: Introduction

Chapter 3: Characterization of compact Cas9 orthologs

Chapter 4: Nme2Cas9 as a compact, hyper-accurate Cas9 with a dinucleotide PAM

ABBREVIATIONS

Acr: Anti-CRISPR

Cas: CRISPR associated

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

crRNA: CRISPR RNA

DMD: Duchenne muscular dystrophy

gDNA: genomic DNA

MGE: Mobile genetic element

NLS: Nuclear localization signal

Indels: Insertions/deletion

DSB: Double-stranded break

HDR: Homology directed repair

RNP: Ribonucleoprotein

ORF: Open reading rrame

ALS: Amyotrophic lateral sclerosis

PAM: Protospacer adjacent motif

PID: PAM interacting domain

sgRNA: Single guide RNA

tracrRNA: Trans-activating crRNA

PUBLISHED WORKS

The following publications appear as whole or in part in this thesis:

Wei Sun, Jing Yang, Zhi Cheng, Nadia Amrani, Chao Liu, Kangkang Wang, Raed Ibraheim, **Alireza Edraki**, Xue Huang, Min Wang, Jiuyu Wang, Liang Liu, Gang Sheng, Erik J. Sontheimer and Yanli Wang. Structures of Neisseria meningitidis Cas9 Complexes in Catalytically Poised and Anti-CRISPR-Inhibited States. *Molecular Cell* (2019)

Bianca Garcia, Jooyoung Lee, **Alireza Edraki**, Yurima Hidalgo-Reyes, Steven Erwood, Aamir Mir, Chantel Trost, Uri Seroussi, Sabrina Y. Stanley, Ronald D. Cohn, Julie M. Claycomb, Erik J. Sontheimer, Karen L. Maxwell, Alan R. Davidson. Potent Inhibition of Cas9 Homologs Used for Genome Editing by a single Anti-CRISPR Protein. *Cell Reports* (2019).

Alireza Edraki, Aamir Mir, Raed Ibraheim, Ildar Gainetdinov, Chunqing Song, Wen Xue and Erik J. Sontheimer. A Compact, high-accuracy Cas9 with a Dinucleotide PAM for *in vivo* Genome Editing. *Molecular Cell* (2019)

Jooyoung Lee, Aamir Mir, **Alireza Edraki**, Bianca Garcia, Nadia Amrani, Hannah E. Lou, Ildar Gainetdinov, April Pawluk, Raed Ibraheim, Xin D. Gao, Pengpeng Liu, Alan R. Davidson, Karen L. Maxwell and Erik J. Sontheimer. Potent Cas9 inhibition in bacterial and human cells by new anti-CRISPR protein families. *mbio* (2018)

Aamir Mir, **Alireza Edraki**, Jooyoung Lee, Erik J. Sontheimer. Type II-C CRISPR-Cas9 Biology, Mechanism and Application. *ACS Chemical Biology* (2017)

Lucas B. Harrington, Kevin W. Doxzen, Enbo Ma, Jun-Jie Liu, Gavin J. Knott, **Alireza Edraki**, Bianca Garcia, Nadia Amrani, Janice S. Chen, Joshua C. Cofsky, Philip J. Kranzusch, Erik J. Sontheimer, Alan R. Davidson, Karen L. Maxwell, Jennifer A. Doudna. A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell* (2017)

April Pawluk, Nadia Amrani, Yan Zhang, Bianca Garcia, Yurima Hidalgo-Reyes, Jooyoung Lee, **Alireza Edraki**, Megha Shah, Erik J. Sontheimer, Karen L. Maxwell, Alan R. Davidson. Naturally occurring off-switches for CRISPR-Cas9, *Cell* (2016)

The following publication that was contributed but is not included in this thesis:

Nadia Amrani, Xin D. Gao, Pengpeng Liu, **Alireza Edraki**, Aamir Mir, Raed Ibraheim, Ankit Gupta, Kanae E. Sasaki, Tong Wu, Paul D. Donohoue, Alexander H. Settle, Alexandra M. Lied, Kyle McGovern, Chris K. Fuller, Peter Cameron, Thomas G. Fazzio, Lihua Julie Zhu, Scot A. Wolfe and Erik J. Sontheimer. NmeCas9 is an intrinsically high-fidelity genome editing platform. *Genome Biology* (2018)

Alireza Edraki and Erik J. Sontheimer. CRISPRs from scratch. *Nature Microbiology* (2018)

Krishna S. Ghanta, Gregory Dokshin, Aamir Mir, **Alireza Edraki**, Hassan Gneid, Jonathon K. Watts, Erik J. Sontheimer and Craig C. Mello. Highly efficient homology-directed repair with chimeric DNA donors. *BioRxiv* (2018)

PREFACE AND COPYRIGHT INFORMATION

Chapter 1

This chapter introduces CRISPR-Cas adaptive immunity and its applications in genome Engineering.

1) Parts of this chapter are taken from our review in *ACS Chemical Biology*
With permission from the American Chemical Society (Copyright 2017, American Chemical Society)

Aamir Mir, **Alireza Edraki**, Jooyoung Lee, Erik J. Sontheimer. Type II-C CRISPR-Cas9 Biology, Mechanism and Application. *ACS Chemical Biology* (2017)

1) Figure 1.1 is adapted from the following article with permission
(license # 4696730352088)
Chen, J.S and Doudna, J. D. (2017) The chemistry of Cas9 and its CRISPR colleagues, *nature reviews chemistry,* 10.1038/s41570-017-0078

2) Figure 1.2 is adapted from the following article under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/)
Koonin, E. V. and Makarova, K. S. (2019) 'Origins and evolution of CRISPR-Cas systems', *Philos Trans R Soc Lond B Biol Sci*

3) Figure 1.3 is adapted from the following article under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/)
Lino, C. A., Harper, J. C., Carney, J. P. and Timlin, J. A. (2018) 'Delivering CRISPR: a review of the challenges and approaches', *Drug Deliv,* 25(1), pp. 1234-1257.

4) Figure 1.4 is adapted from the following article under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/)
Ding, Y., Li, H., Chen, L. L. and Xie, K. (2016) 'Recent Advances in Genome Editing Using CRISPR/Cas9', *Front Plant Sci,* 7, pp. 703.

5) Figure 1.6 is adapted from the following article with permission
(license # 4693120374924)
Germini, D., Tsfasman, T., Zakharova, V. V., Sjakste, N., Lipinski, M. and Vassetzky, Y. (2018) 'A Comparison of Techniques to Evaluate the Effectiveness of Genome Editing', *Trends Biotechnol,* 36(2), pp. 147-159.

Chapter 2

Chapter 2 describes the discovery and characterization of phage-encoded anti-CRISPR proteins. This was a collaborative project and parts of it are published as follows.

1) Chapter 2.2.1 is a collaboration between members of the Sontheimer, Davidson and Maxwell labs, particularly April Pawluk, Nadia Amrani, Yan Zhang, Bianca Garcia and me. It is published (see below) and requires no permissions for reprints for authors.

April Pawluk, Nadia Amrani, Yan Zhang, Bianca Garcia, Yurima Hidalgo-Reyes, Jooyoung Lee, **Alireza Edraki**, Megha Shah, Erik J. Sontheimer, Karen L. Maxwell, Alan R. Davidson. Naturally occurring off-switches for CRISPR-Cas9, *Cell* (2016)

2) Chapters 2.2.2 and 2.2.4 is a collaboration between members of the Sontheimer, Davidson and Maxwell labs. The experiments in figure 2.3 was performed by Aamir Mir, Hannah Lou, Nadia Amrani and Jooyoung Lee. Parts of the chapter are published (see below) under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/)

Jooyoung Lee, Aamir Mir, **Alireza Edraki**, Bianca Garcia, Nadia Amrani, Hannah E. Lou, Ildar Gainetdinov, April Pawluk, Raed Ibraheim, Xin D. Gao, Pengpeng Liu, Alan R. Davidson, Karen L. Maxwell and Erik J. Sontheimer. Potent Cas9 inhibition in bacterial and human cells by new anti-CRISPR protein families. *mbio* (2018)

3) Chapter 2.2.3 is a collaboration between the Sontheimer lab, Doudna lab and Maxwell lab. The chapter includes parts of the following publication. It is published (see below) and requires no permissions for reprints for authors.

Lucas B. Harrington, Kevin W. Doxzen, Enbo Ma, Jun-Jie Liu, Gavin J. Knott, **Alireza Edraki**, Bianca Garcia, Nadia Amrani, Janice S. Chen, Joshua C. Cofsky, Philip J. Kranzusch, Erik J. Sontheimer, Alan R. Davidson, Karen L. Maxwell, Jennifer A. Doudna. A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell* (2017)

Chapter 3

Chapter 3 relates to the discovery of compact Cas9 orthologs for genome editing. It is a highly collaborative project with several labs involved. The data in table 3.2 are from Bianca Garcia and the other data were acquired in collaboration with Jooyoung Lee, Raed Ibraheim and Aamir Mir. Parts of the chapter are published as follows that require no permission for authors.

Bianca Garcia, Jooyoung Lee, **Alireza Edraki**, Yurima Hidalgo-Reyes, Steven Erwood, Aamir Mir, Chantel Trost, Uri Seroussi, Sabrina Y. Stanley, Ronald D. Cohn, Julie M. Claycomb, Erik J. Sontheimer, Karen L. Maxwell, Alan R.

Davidson. Potent Inhibition of Cas9 Homologs Used for Genome Editing by a single Anti-CRISPR Protein. *Cell Reports* (2019).

Jooyoung Lee, Aamir Mir, **Alireza Edraki**, Bianca Garcia, Nadia Amrani, Hannah E. Lou, Ildar Gainetdinov, April Pawluk, Raed Ibraheim, Xin D. Gao, Pengpeng Liu, Alan R. Davidson, Karen L. Maxwell and Erik J. Sontheimer. Potent Cas9 inhibition in bacterial and human cells by new anti-CRISPR protein families. *mbio* (2018)

Chapter 4

This chapter describes the discovery and characterization of Nme2Cas9. This chapter is the culmination of a collaborative effort between Sontheimer, Rivera, xue and Wang labs. Ildar Gainetdinov performed bioinformatics analyses for PAM identification experiments. Except for chapter 4.2.7, all other sections are published in the following two publications and require no permission for authors for reprints.

Wei Sun, Jing Yang, Zhi Cheng, Nadia Amrani, Chao Liu, Kangkang Wang, Raed Ibraheim, **Alireza Edraki**, Xue Huang, Min Wang, Jiuyu Wang, Liang Liu, Gang Sheng, Erik J. Sontheimer and Yanli Wang. Structures of Neisseria meningitidis Cas9 Complexes in Catalytically Poised and Anti-CRISPR-Inhibited States. *Molecular Cell* (2019)

Alireza Edraki, Aamir Mir, Raed Ibraheim, Ildar Gainetdinov, Chunqing Song, Wen Xue and Erik J. Sontheimer. A Compact, high-accuracy Cas9 with a Dinucleotide PAM for *in vivo* Genome Editing. *Molecular Cell* (2019)

Chapter 5

In Chapter 5, we present preliminary data on using CRISPR-Cas9 genome editing as a therapeutic approach for ALS. This is a collaboration between Sontheimer and Mueller labs, particularly with Raed Ibraheim and Karin Meijboom. All data in this chapter are unpublished

CHAPTER 1: Introduction

1.1 Introduction to CRISPR-Cas

1.1.1 The biology and diversity of CRISPR-Cas

Bacteriophages, or viruses that infect bacteria, are the most common biological entity on the planet, with an estimated 10^{31} bacteriophages on earth. As a response, bacteria have evolved dozens of defense systems to combat phage infection. One category of defense system is CRISPR-Cas [Clustered, regularly interspaced, short palindromic repeats along with CRISPR-associated (Cas) proteins], which constitutes bacterial and archaeal adaptive immune pathways against phages and other mobile genetic elements (MGEs) (Barrangou *et al.*, 2007; Marraffini and Sontheimer, 2008; Brouns *et al.*, 2008). CRISPR is present in most archaea and about half of all bacteria and provides immunity against MGEs by targeting and degrading their associated nucleic acids.

Thousands of CRISPR-Cas systems have been identified thus far. Despite tremendous diversity, the architectures of CRISPR loci share many similarities. CRISPR immunity consists of an adaptation phase and an interference phase. The adaptation phase provides an immune genetic memory of previous infections in the form of DNA. Next, the DNA is transcribed, processed into mature CRISPR RNAs (crRNAs, also called crRNA processing) and serves as a guide for Cas endonucleases to find and cleave future invaders, a process known as interference (Figure 1.1). A

typical CRISPR locus consists of a CRISPR array flanked by Cas genes. The CRISPR array consists of a series of identical, short (~25-60) repeat sequences (Shmakov *et al.*, 2017b) interspaced by MGE-derived spacer sequences of similar length, which are complimentary to the target and form the basis of CRISPR immunity. Cas proteins are involved in adaptation, crRNA processing, and interference.

Numerous Cas proteins have been identified to date and can be categorized into adaptation, processing and effector modules based on their functions (Koonin and Makarova, 2009; Makarova *et al.*, 2015; Koonin and Makarova, 2019). Only Cas1 and Cas2, which function in CRISPR adaptation, are nearly universal to all CRISPR-Cas systems. In contrast, CRISPR effector modules (i.e. the interference complex) are extremely diverse and form the basis of the current classification scheme of CRISPR systems, which broadly groups them into two classes (class I and class II) comprising six types (Shmakov *et al.*, 2017a). Class I systems employ multi-protein effector complexes, whereas Class II CRISPR systems use a single effector protein for interference (Figure 1.2A).

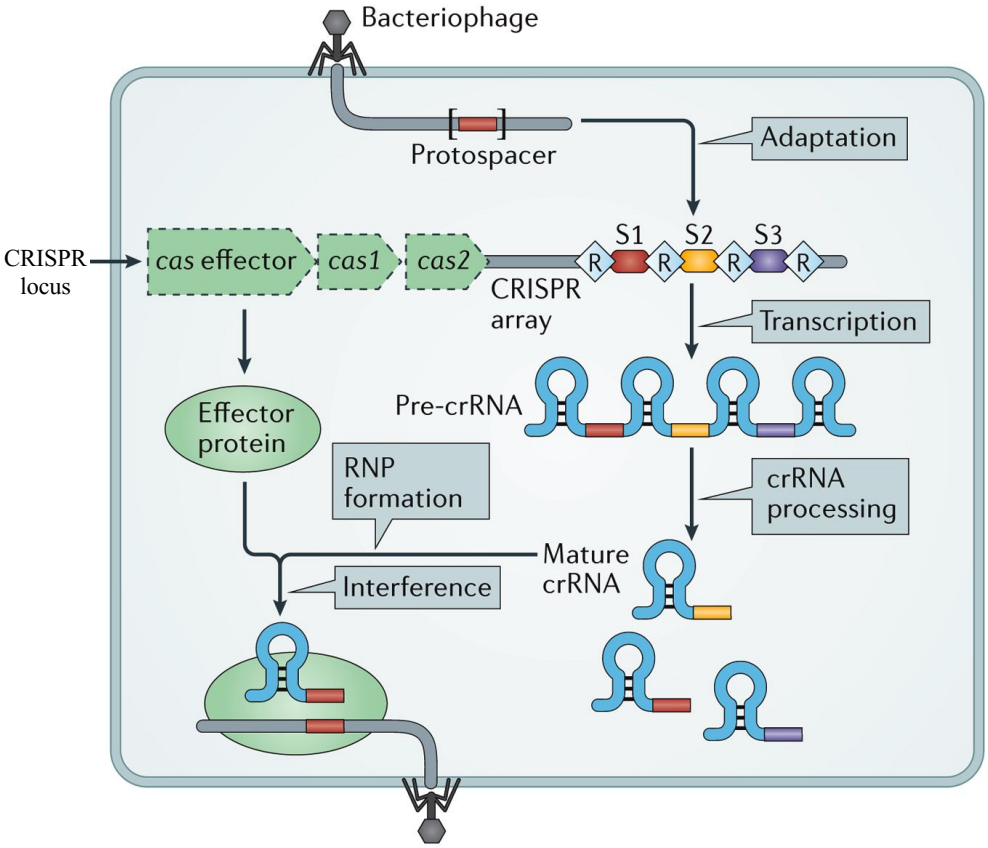


Figure 1.1 CRISPR as an adaptive immune pathway
A typical CRISPR locus consists of a CRISPR array and several Cas proteins. CRISPR adaptive immunity begins with the acquisition of a spacer from invading MGEs (adaptation) and integrating them into the CRISPR array. CRISPR loci give rise to crRNAs (crRNA processing), which guide the effector module Cas proteins to their protospacer targets (interference).

As part of the constant arms race between phages and bacteria, phages have evolved many counter measures against CRISPR including mutations and anti-CRISPR proteins (Acrs). Acrs are small proteins that bind to CRISPR effectors and render them inactive, disarming CRISPR defense and allowing phage propagation. Acrs are highly diverse and have been reported to inhibit CRISPR effectors via distinct mechanisms. In chapter 2, we will describe some of these Acrs that inhibit CRISPR-Cas9 and provide mechanistic insight into their inhibitory function.

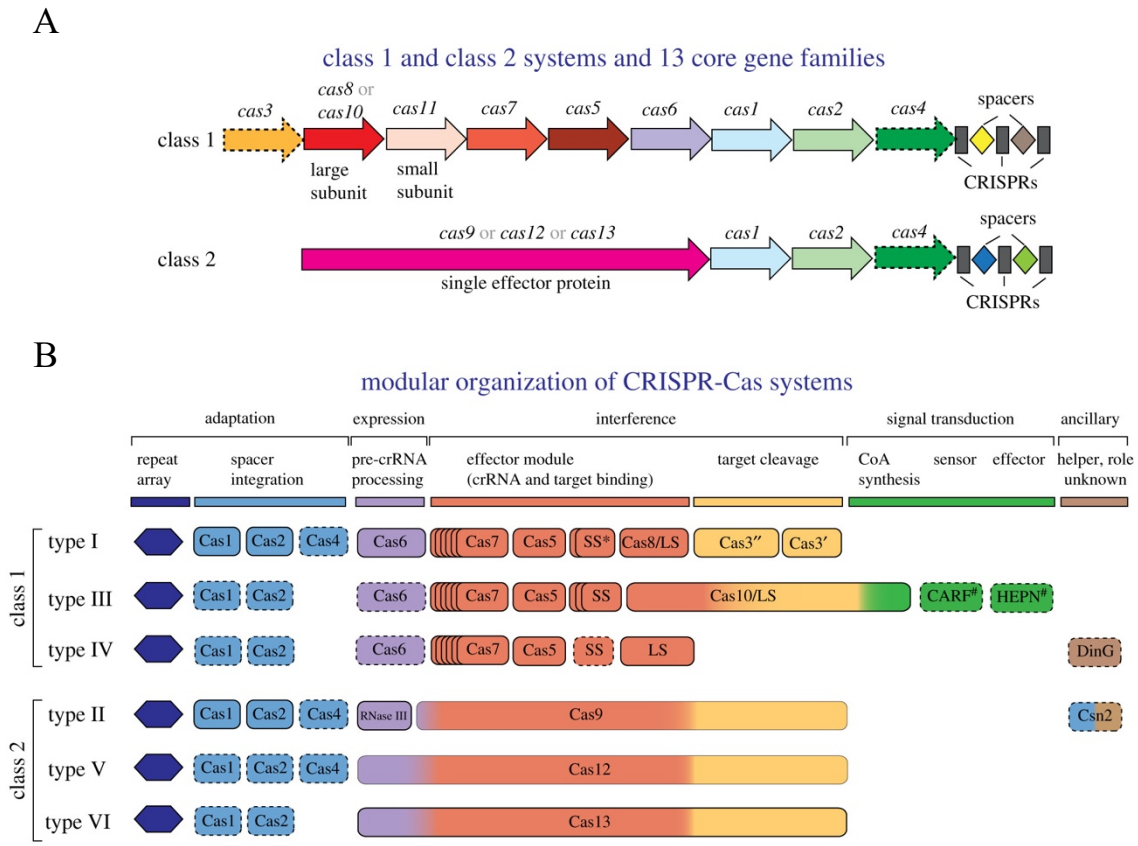


Figure 1.2 Classification and diversity of CRISPR systems
(A) The basis for class 1 and class 2 CRISPR systems. Class 1 systems are characterized by the presence of multiple subunits in their effector complex, while class 2 systems have a single, RNA-guided nuclease for interference. Cas1 and Cas2 are almost universal and are part of the adaptation complex.
(B) There are three types of CRISPR systems in each class. Dispensable genes (not found in all systems) in each type are indicated by a dashed outline. Adapted from Koonin & Makarova [(Koonin and Makarova, 2019) with permission (see preface).

1.1.2 A brief history of CRISPR

Like most discoveries in science, CRISPR has a rich history with hundreds of scientists involved, building upon the works of the previous scientists to culminate to what today has become one of the most important advances in science. Here, I will outline some of the key CRISPR discoveries, which is by no means a comprehensive history of CRISPR.

In 1987, Ishino and colleagues were working on the isozyme conversion of alkalinephosphatase in *E. coli* when they came across a series of identical repeats flanking their gene of interest. In their discussion, they wrote that "an unusual structure was found in the 3'end flanking region of *iap* [their gene of interest]. Five highly homologous sequences of 29 nucleotides were arranged as direct repeats with 32 nucleotides as spacing" (Ishino *et al.*, 1987). About three decades later, we now know that this was the first reference to CRISPR.

The term CRISPR was coined in 2002 by Ruud Jansen and colleagues, who showed that these loci were widespread in bacteria and archaea but are absent in eukaryotes (Jansen *et al.*, 2002). These repeats had already captured the attention of Francisco Mojica, who studied these repeats and showed that they derive from foreign genetic elements and may be involved in defense against MGEs (Mojica *et al.*, 2005). In fact, Mojica had initially dubbed these loci as short regularly spaced repeats (SRSRs) but later on recommended that Jansen *et al.* emphasize the palindromic nature of repeats, which resulted the arguably catchier acronym, CRISPR. This was followed shortly after by the discovery of Cas proteins.

Speculations regarding the nature of CRISPR-Cas continued until 2007, when the first evidence of CRISPR's function was shown by Barrangou *et al.* They demonstrated that bacteria exposed to phages could use CRISPR to combat the attack (Barrangou *et al.*, 2007). Several mechanistic studies from 2008 to 2012 culminated in two papers by Jinek *et al*. and Gasiunas *et al.* demonstrating that Cas9 (a type II CRISPR-associated endonuclease) is an RNA-guided endonuclease. These biochemical studies paved the way for heterologously expressing this bacterial protein to be used for genome engineering (Jinek *et al.*, 2012; Gasiunas *et al.*, 2012). The CRISPR craze kickstarted with several papers in 2013 establishing Cas9 as a genome engineering tool and Cas9 has been widely used ever since (Cong et al., (Esvelt *et al.*, 2013; Mali *et al.*, 2013; Hou *et al.*, 2013; Cong and Zhang, 2015).

1.2 Type II CRISPR-Cas systems

The most common and well-studied Class II systems are from type II, which uses the Cas9 protein as its effector and will be the primary focus of this thesis. Type II CRISPR systems are subdivided into three subtypes based on the degree of homology between Cas9 proteins, and on the presence or absence of other Cas proteins besides Cas1, Cas2 and Cas9 (Shmakov *et al.*, 2017a). Type II-A systems have an additional protein named Csn2, type II-Bs are distinguished by the presence of Cas4, and type II-Cs are characterized by the absence of both Cas4 and Csn2. Recently, additional variants of type II-C CRISPR systems have also been identified in archaea that share similarity with type II-C Cas9s but also contain Cas4 (Burstein *et al.*, 2017). Subtypes II-A, -B and -C comprise ~55%, ~3% and ~41% of type II systems identified in public sequence database (Shmakov *et al.*, 2017a). Like other CRISPR systems, the mechanism of CRISPR immunity is still divided into three steps (figure 1.1), adaptation, crRNA processing and interference:

1) Adaptation: Adaptation is the first step in CRISPR immunity. Most of our knowledge of adaptation comes from the type I-E system of *Escherichia coli (E. coli)*, and the type II-A systems of *S. pyogenes* and *Streptococcus thermophilus* (Jackson *et al.*, 2017; Sternberg *et al.*, 2016). In most systems studied to date, Cas1 and Cas2 form a complex to capture small fragments of DNA from the MGE and integrate them into the CRISPR array. In type II-A systems, Cas9 was also required for successful adaptation, in part to enforce PAM specificity of the DNA fragments being captured

(Heler *et al.*, 2015). These spacers along with the repeats are then transcribed in the second step in CRISPR immunity, also known as crRNA biogenesis.

2) CRISPR RNA Biogenesis: In most CRISPR-Cas systems, a single pre-crRNA transcript spanning the CRISPR locus is initiated from the leader region, and the pre-crRNA must be processed to yield individual crRNAs (Charpentier *et al.*, 2015; Deltcheva *et al.*, 2011). This is also the case for type II-A and II-B systems studied to date. In contrast, many type II-C arrays possess internal, independent promoters embedded in each repeat sequence, yielding nested sets of pre-crRNAs of varying length (Zhang *et al.*, 2013).

In addition to a crRNA, type II systems require a trans-activating crRNA (tracrRNAs) (Figure 1.3). tracrRNAs include a complementary region that hybridizes with repeat sequences within pre-crRNAs (Deltcheva *et al.*, 2011). Processing involves endonucleolytic cleavage of both strands of the pre-crRNA:tracrRNA duplex by the host factor RNase III (Charpentier *et al.*, 2015).

3) Interference: Cas9 is the sole protein responsible for interference in type II systems. Cas9 forms a ribonucleoprotein (RNP) complex with crRNA and a tracrRNA, and this complex cleaves DNA in a crRNA-guided manner (Figure 1.3). Self-targeting is avoided in all type II systems by requiring a protospacer adjacent motif (PAM) sequence for target engagement and cleavage. PAM recognition by Cas9's PAM interacting domain (PIDs) is required before the initiation of dsDNA cleavage (Sternberg *et al.*, 2015). In the context of adaptive immunity, PAMs are used to

distinguish self vs. non-self DNA. i.e. since the complementary sequence is present in both the CRISPR array and the MGE, effector proteins use PAMs to only target the MGE. In the case of the spacer in the CRISPR array, the PAM sequence gets removed prior to spacer insertion by the adaptation machinery.

Despite the tremendous diversity of Cas9s, they all share a similar domain structure (figure 1.4). First, The N-termini of Cas9s start with the catalytic RuvC-I domain, which is responsible for the cleavage of the displaced (crRNA-noncomplementary) strand of the protospacer. The arginine-rich bridge helix (BH) follows the RuvC-I domain and has previously been implicated in guide RNA recognition, especially in the region that pairs with the target DNA nucleotides that are nearest the PAM (the "seed" region). The α-helical recognition (REC) lobe is composed of several REC domains that are involved in the recognition of the target protospacer. Next, the HNH (His-Asn-His) endonuclease domain, which cleaves the crRNA-complementary DNA strand, is sandwiched between the RuvC-II and RuvC-III domains. Lastly, the highly divergent PAM-interacting domain (PID) is found at the C terminus, which serves to recognize the diverse PAMs found in type II systems (Shmakov *et al.*, 2017a; Mir *et al.*, 2018b).

Once the RNA is processed and loaded onto Cas9, there are conformational changes that result in the ability of Cas9 to recognize the PAM in the target dsDNA. Specifically, studies on SpyCas9 have shown several conformational changes within the REC and HNH domains upon sgRNA loading and DNA binding (Sternberg *et al.*,

2015). The cleavage of target strand by the HNH domain is immediately followed by the cleavage of the non-target strand by the RuvC domain. The DSBs thwart the propagation of MGEs in a sequence-specific manner.

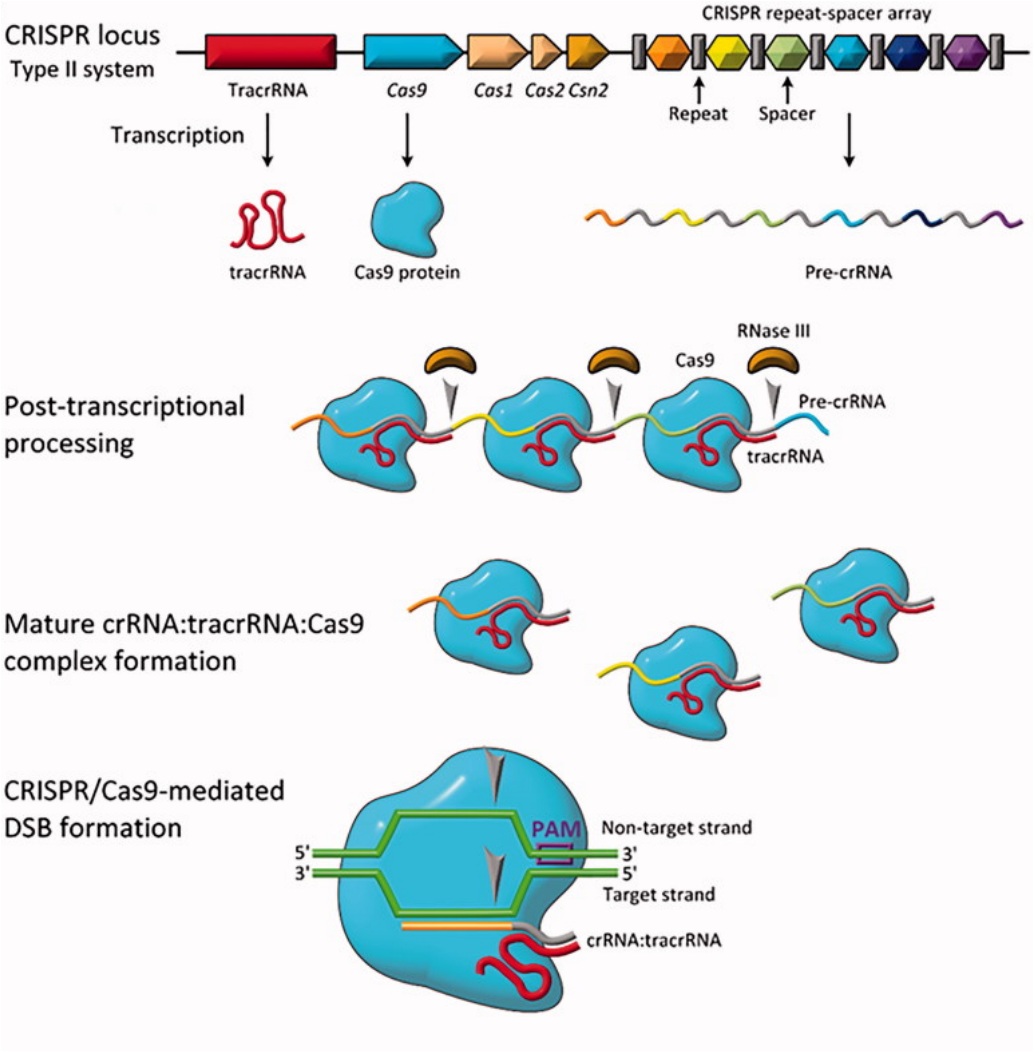


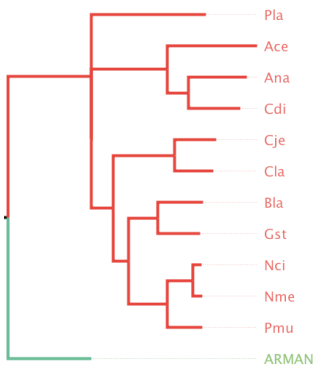
Figure 1.3: The biology of type II CRISPR systems
A typical type II-A system is shown. These crRNA guides are initially transcribed into longer pre-crRNAs that are then processed into their mature forms. The spacers within the CRISPR locus derive from previously encountered MGEs and are incorporated by a process known as adaptation.

Among type II CRISPR subtypes, type II-C is the simplest and often possesses relatively smaller Cas9s (Figure 1.4A). The smaller size of these orthologs makes them suitable for Adeno-associated virus (AAV) delivery, which has limited packaging capacity (see section 1.4.1) (Lino *et al.*, 2018). Of the ~4,000 Cas9 orthologs currently in the NCBI database, ~1,500 are type II-C (Shmakov *et al.*, 2017a). These type II-C systems are found in diverse bacterial species that grow in extremely different environments, from acidic hot springs in Yellowstone National Park (*Acidothermus cellulolyticus*), to the respiratory tract of pigs (*Pasteurella multocida*), and to waste waters in Thailand (*Tisterilla mobilis*) (Mir *et al.*, 2018b). This environmental breadth likely drives the evolution of the unique and diverse Cas9 orthologs observed. In fact, an analysis of type II-C systems reveals substantial sequence diversity among components of type II-C loci, including their Cas9 orthologs (Chylinski *et al.*, 2014). Specifically, the diversity within type II-C PIDs suggests the recognition of divergent PAMs. This is demonstrated by type II-C Cas9 PAMs defined to date (Figure 1.4B). Interestingly, type II-C Cas9 orthologs in closely related species of Campylobacter (e.g. *C. jejuni* and *C. lari*) can recognize highly dissimilar PAMs (Ran *et al.*, 2015; Kim *et al.*, 2017).

Another peculiar aspect of some type II-C Cas9 orthologs is that their associated PAMs can extend relatively far from the crRNA-complementary sequence (often up to 8 nt). The long PAMs may influence target site choice and off-target effects of type II-C Cas9s, as discussed further below.

A

B



Cas9	PAM	Size (aa)	Validated for mammalian genome editing?	References
Parvibaculum lavamentivorans	NNNCAT	1037	N	Ran et al. (2015)
Acidothermus cellulolyticus	NNNCC	1138	N	Tsui et al. (2017)
Actinomyces naeslundii	?	1101	N	Ma & Harrington et al. (2015)
Corynebacterium diphtheriae	NGG	1084	N	Ran et al. (2015), Ma & Harrington et al. (2015)
Campylobacter jejuni	NNNNRYAC	984	Y	Fonfara et al. (2014), Kim, Koo & Park et al. (2016), Yamada & Watanabe et al. (2017)
Campylobacter lari	NNGGGT	1003	N	Ran et al. (2015)
Brevibacillus laterosporus	NNNNCNDD	1092	Y	Karvelis et al. (2015)
Geobacillus stearothermophilus	NNNNCNAA	1087	Y	Harrington et al. (2017)
Neisseria cinerea	NNNNGTA?	1082	N	Ran et al. (2015)
Neisseria meningitidis	NNNNGNTT	1082	Y	Hou & Zhang et al. (2013), Esvelt & Mali et al. (2013), Fonfara et al. (2014), Lee et al. (2016)
Pasteurella multocida	GNNNCNNA	1056	N	Fonfara et al. (2014)
Archaeal	NGG	950	N	Burstein, Harrington & Strutt et al. (2017)

Figure 1.4 Domain architecture and diversity of type II-C Cas9s
(A) The domain architecture of type II-C Cas9s. The individual domains are colored separately and labeled. (B) A phylogenetic tree of *in vitro* validated type II-C Cas9s. The reported PAMs and the protein sizes of Cas9s are also listed. Cas9s validated for mammalian editing are also marked as Y (Yes) or N (No).

1.3 CRISPR-Cas9 genome engineering

1.3.1 Applications of CRISPR-Cas9 in biomedical sciences

The ability to modify the genome can provide the opportunity to study and treat human disease. Genetic disorders caused by abnormalities in a gene(s) are widespread throughout the world. Online Mendelian Inheritance in Man (OMIM) estimates that over 6500 different diseases are the result of monogenic mutations, often with a single nucleotide difference between patients and healthy individuals. Consequently, genome editing i.e. the ability to make changes to a target DNA of interest can be used to model or treat such diseases. It is therefore not surprising that several genome editing approaches have been devised in the past few decades. Most of these approaches are based on a nuclease that can be programmed to a target of interest. For example, zinc-finger nucleases are engineered by fusing a DNA binding domain to a DNA cleavage domain. DNA cleavage triggers the cellular repair machinery to repair the break, which forms the basis of genome editing applications (see chapter 1.3.3 for mechanistic details of repair outcomes). However, these methods were expensive, time consuming and often inefficient. As a result, the advent of Cas9 genome editing was timely, which will be discussed below.

CRISPR-Cas has revolutionized biomedical sciences and enumerating its uses would encompass hundreds of pages, so here I will just list a handful of its applications. The ability to target sites using RNA-DNA complementarity rather than protein engineering is the transformative feature of Cas9 and other RNA-programmable Cas

proteins. Nuclease Cas9 has been successfully used in model organisms such as *Mus musculus* (mice), *Caenorhadbitis elegans* (round worm), *Escherichia coli* (bacterium), *Saccharomyces cerevisiae* (yeast), *Arabidopsis thaliana* (plant). Applications include creating gene knockouts by simply inducing indels or creating a desired outcome by providing donors for homologous recombination (Doudna and Charpentier, 2014). In addition, Cas9 has been used to generate disease models in human induced pluripotent stem cells (iPSCs), which can be differentiated into the cell type of interest (Musunuru, 2013).

Genome-wide CRISPR screens have been powerful in identifying disease-associated protein functions in a systemic manner. Most CRISPR screens use thousands of guide RNAs to knock out genes across millions of cells while applying a selective pressure, and then identify protospacer sequences that are either enriched or depleted in the selected cell population relative to control. For example, Wang *et al* used a library of 73,000 sgRNAs to generate knockout collections and performed screens to identify essential genes in DNA mismatch repair pathways (Wang *et al.*, 2015). While CRISPR screens require detectable phenotypic (often survival) outcomes, they have proven powerful in untangling complex pathways and uncovering important genes in cancer and other diseases.

In addition to wildtype Cas9 (wtCas9), catalytically dead Cas9 (dCas9) has been used to deliver other enzymes to a locus of interest. Fusion of dCas9 to transcriptional activators (CRISPRa) and repressors (CRISPR interference, CRISPRi) has resulted in

fine-tuned, reversible transcriptional regulation in bacteria and eukaryotes (Qi *et al.*, 2013; Larson *et al.*, 2013). dCas9 has also been fused to fluorescent proteins for visualization of genetic loci of interest (Chen *et al.*, 2013). The advent of base editing is another outcome of the CRISPR revolution. Base editors consist of Cas9 fused to a cytidine deaminase, guided to a locus of interest resulting in an ultimate C-G → T-A conversion without a DSB (Komor *et al.*, 2016). More recently, an RNA adenosine deaminase was evolved to deaminate deoxyadenosine, and was fused to dCas9 for A-T→G-C conversion (Gaudelli *et al.*, 2017). In summation, Cas9-based applications have helped scientists throughout the world to uncover basic molecular mechanisms of healthy and diseased cells.

1.3.2 Clinical applications of CRISPR-Cas9

CRISPR-Cas9 has already been used for therapeutic and diagnostic purposes. Proof-of-concept experiments have demonstrated that Cas9 editing in human embryos can correct deleterious mutations. For example, scientists used Cas9 to correct MYBPC3 mutations (causes hypertrophic cardiomyopathy) and similar efforts were done on the β-globin gene (Ma *et al.*, 2017; Liang *et al.*, 2015). There is also ongoing work on treating genetic defects in adults. Due to the outstanding delivery challenges (see chapter 1.4), most of the current trials rely on *ex vivo* approaches. Such approaches consist of editing patient cells outside of the body and reintroducing the cells with corrected gene(s). For example, there are currently multiple clinical trials for *ex vivo* therapeutics of β thalassemia and sickle cell disease, two life-threatening afflictions (clinicaltrials.gov, 2019).

The rapid implementation of CRISPR technologies has generated controversy regarding safety and ethics. In 2016, scientists in China injected CRISPR-edited cells into the lungs of a lung cancer patient to arm T-cells to better fight cancer (Cyranoski, 2016). In 2018, the birth of the first ever genetically edited humans was reported, leading to considerable controversy and ethical debate. Although controversial, these efforts and studies highlight the power of CRISPR-Cas9 genome editing.

1.3.3 The mechanism of CRISPR-Cas9 genome engineering

The RNA programmability of Cas9 has made it a powerful genome editing tool in biotechnology and medicine (Cho *et al.*, 2013; Esvelt *et al.*, 2013; Mali *et al.*, 2013; Cong and Zhang, 2015). In engineered systems (e.g. for genome editing), the crRNA and tracrRNA can be fused into a single-guide RNA (sgRNA). Cas9 is guided to the locus of interest to induce a double-stranded break (DSB, when both strands of the DNA double-helix are cut). DSBs are inherently dangerous for the cell, as they could lead to rearrangements and loss of genetic information. Cells have evolved repair pathways to deal with DSBs and taking advantage of these repair pathways enables editing of specific loci.

Since DSBs threaten the integrity of the genome, Cas9-mediated cleavage triggers a repair response (Figure 1.5). In most mammalian cells, DSBs are frequently repaired via an imprecise process called non-homologous end-joining (NHEJ). Sometimes, the NHEJ machinery ligates the blunt ends of the Cas9-cleaved DNA, restoring the original sequence that is susceptible to another round of cleavage. On the other hand, sometimes NHEJ inserts or deletes a few bases prior to ligation. This results in nucleotide insertions/deletions (indels) that inactivate the target gene (Lino *et al.*, 2018). Creating indels in a gene of interest is powerful, since we can study gene function and in the case of a disease-causing gene, knockout its function.

An alternative repair pathway is homology-directed repair (HDR), which requires that a donor DNA is provided and is often less efficient than NHEJ (Figure

1.5). The template DNA can be either from an uncut allele in the genome, or exogenously supplied dsDNA or ssDNA donors. The HDR pathway is crucial in repairing disease-causing mutations that need to be precisely repaired and NHEJ knockout is not sufficient. Consequently, the naturally low rates of HDR is a hindrance to some therapeutic applications. To this end, mutations in Cas9 nickases, in which the active sites of either the HNH or RuvC domain is mutated, have been used to improve HDR and to improve genome editing specificity via DSB induction by dual nickases (Ran *et al.*, 2015) (Mali *et al.*, 2013).

Theoretically, any sequence in the genome that is adjacent to a PAM can be targeted. However, editing varies considerably from site to site, depending on sgRNA expression, DNA accessibility and other factors. Furthermore, some Cas9 orthologs have the propensity to cleave near-cognate sites (known as “off-target” events, discussed further in 1.4.3 and 1.5) which must be taken into account during target selection.

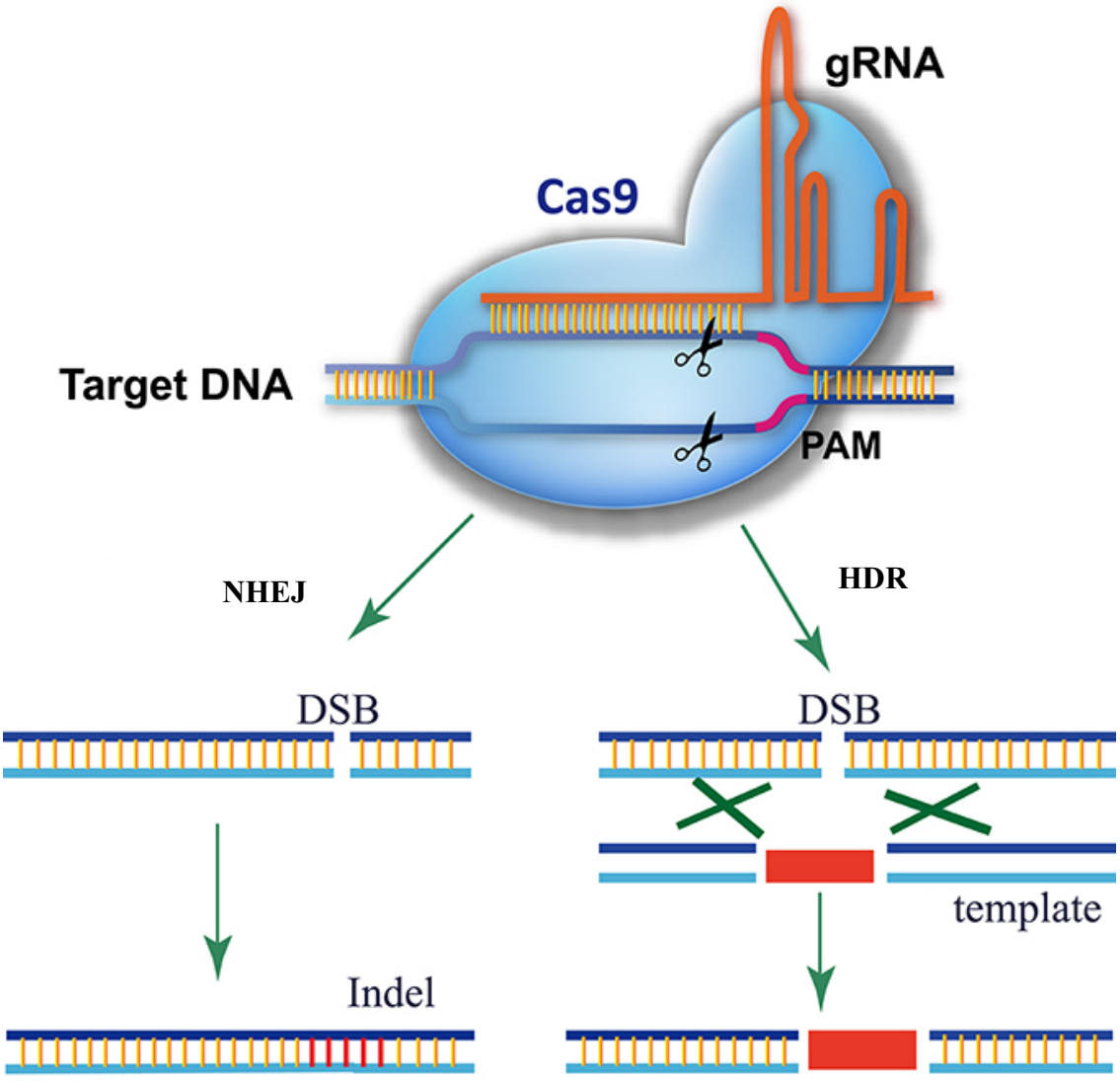


Figure 1.5 DNA repair outcomes in mammalian cells
Following Cas9-mediated cleavage, most mammalian cells repair the double-strand break via NHEJ, which could lead to Indels(left). Alternatively, cells can repair the break via HDR if a donor is provided (right). Figure adapted from (Ding *et al.*, 2016) with permission.

1.3.4 Methods to quantify CRISPR-Cas9 genome editing

It is important to accurately quantify genome editing outcomes and assess the spectrum of indels (in the case of NHEJ) or the rate of HDR at genomic loci. For example, if gene knockout is desired, it is important to measure what percent of the cells have indels and what percent of the indels result in in-frame vs out-of-frame mutations. Different methods to detect indels have been devised and these can be broadly divided into sequencing-based and denaturation-based. In this thesis, we employ three of the most established methods to quantify genome editing events, illustrated in Figure 1.6.

The T7 endonuclease assay, also known as the T7E1 or the surveyor assay, is a denaturation-based method that takes advantage of the ability of T7E1 to cleave mismatches in DNA heteroduplexes. It consists of amplifying the target locus of interest via polymerase chain reaction (PCR), followed by denaturation and reannealing to create heteroduplexes. Next, the PCR fragments are subjected to T7E1 cleavage and the products can be visualized on an agarose gel. This is a rapid and cost-effective approach to measure NHEJ events from a population of cells. However, it is not very sensitive; it underestimates events due to the inefficiency of T7E1 in cleaving single nucleotide mismatches, and it does not reveal the spectrum of indels (Germini *et al.*, 2018).

We also used two sequencing-based methods in this thesis to more accurately measure indels. Targeted deep sequencing is one of the most accurate method to quantify genome editing events. It consists of two rounds of PCR of the target locus,

one to amplify the segment and the second to add the appropriate tags and barcodes for the high-throughput sequencer to enable multiplexing. Targeted deep sequencing has the added advantage of reliably revealing the nature of indels (the length as well as the sequence) and can sensitively detect indels as low as 0.1%, depending on the DNA and sequencing quality. The disadvantages of deep sequencing include long processing time, high costs and requiring extended analysis.

Tracking of Indels by Decomposition (TIDE) is an algorithm devised by Brinkman et al. that uses trace decomposition from Sanger sequencing to measure indels (Brinkman *et al.*, 2014). TIDE has a lower detection limit of ~1-2%, and highly correlates with deep sequencing data (Germini *et al.*, 2018). Relying on Sanger sequencing provides a rapid and relatively cheap method of quantifying indels, although the accuracy heavily relies on DNA fragment and sequencing quality.

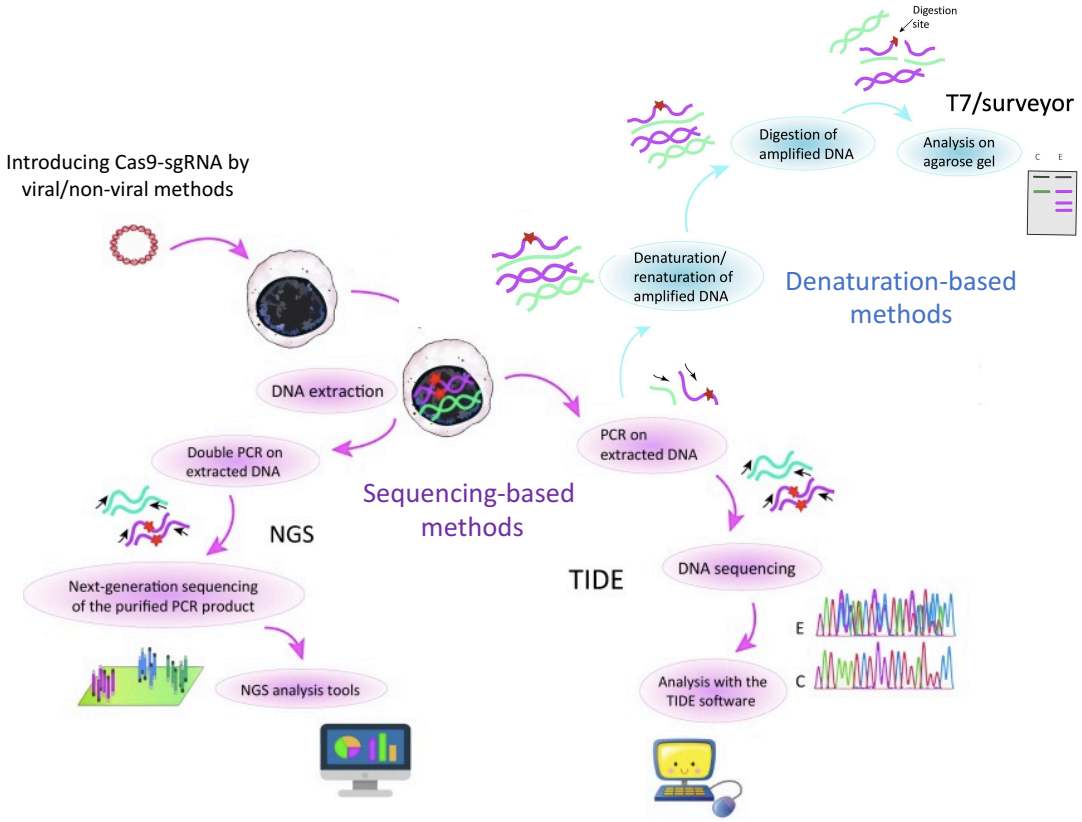


Figure 1.6 Methods used in this thesis to quantify genome editing by Cas9
Two sequencing-based methods (shown in purple) are used, namely targeted deep sequencing and TIDE. Targeted deep sequencing requires two rounds of PCR followed by sequencing using a next generation sequencer. TIDE is an algorithm to decompose Sanger sequencing trace files to estimate indels. We also used the T7E1 assay (shown in blue), a rapid method to quantify indels by taking advantage of T7 endonuclease's ability to cleaved mismatched DNA. Figure adapted from (Germini *et al.*, 2018).

1.4 Methods for *in vivo* Cas9 delivery

Before Cas9 can cleave its target, it needs to be delivered to the target cell of interest. Current delivery methods of Cas9 to cells can be divided broadly into cell culture/*ex vivo* approaches and *in vivo* delivery. For the purpose of this thesis, *ex vivo* refers to editing cells outside of the organism and reintroducing them. *In vivo* refers to editing that occurs inside of an organism. All experiments in cell culture are so indicated.

In cell culture and *ex vivo*, there are many viral and non-viral ways to efficiently deliver Cas9-sgRNA into cells. Delivery methods such as chemical transfection of plasmids, lentiviral transduction, and ribonucleoprotein (RNP) electroporation have been used to deliver Cas9-sgRNA into various cell types (Lino *et al.*, 2018). Most non-viral approaches rely on the temporary disruption of the cell membrane to introduce plasmids/mRNAs/RNPs. Viral delivery often consists of the recognition of surface receptors that lead to uptake. In most cases, Cas9 is fused to a nuclear localization signal (NLS) that mediates intranuclear uptake. While efficient editing in certain primary cells still remains a challenge, delivering Cas9 in cell culture is relatively straightforward compared to *in vivo*.

In vivo delivery (particularly in humans) is an even a greater challenge due to need for high safety and precision. Some of the potential approaches include Adeno-associated virus (AAV), lentivirus, lipid nanoparticles/liposomes, DNA nanoclew, and gold nanoparticles (Liu *et al.*, 2017). While every approach has unique advantages and

disadvantages, some have stood the test of time more than others. For example, lentiviruses may result in genome rearrangements and are highly immunogenic. There has been tremendous progress in lipid and nanoparticle delivery methods, but they still suffer from low stability, poor distribution and lysosomal degradation (Liu *et al.*, 2017; Lino *et al.*, 2018). While all of these approaches have been used to deliver Cas9, AAV delivery has been the gold standard for *in vivo* Cas9 delivery since 2013.

1.4.1 AAV delivery of Cas9

AAV was discovered accidentally during electron microscopy screening of adenovirus preparations in the 1960s (Keeler and Flotte, 2019). Ever since, AAV delivery has been one of the most promising approaches, with decades of research on its safety and efficacy. AAV is a small virus that infects primates without causing any diseases. It has a packaging capacity of ~4700 bp and can be engineered to persist extrachromosomally with minimal integration into the host genome (Lino *et al.*, 2018). AAV vectors can infect both dividing and postmitotic cells, expanding their use to cells like neurons. Furthermore, there are different serotypes of AAV (serotypes differ in the receptors they bind) which enable delivery to the desired target tissue (Lau and Suh, 2017). These advantages have led to dozens of ongoing clinical trials using AAV as the delivery vector of choice, with the first AAV drug approved by the FDA in 2017 (Luxturna) to treat Leber's congenital amaurosis. This was followed by Zolgensma in 2019 to treat Spinal Muscular Atrophy. Taken together, AAV holds tremendous potential as the delivery method of choice for CRISPR-Cas9 therapeutics.

Several proof-of-concept studies have used AAV to deliver Cas9 for therapeutic applications in model organisms. For example, Duchenne muscular dystrophy (DMD) is a devastating disease caused by mutations in the dystrophin gene. AAV-based Cas9 treatment of a mouse model with mutants in the dystrophin gene led to restoration of gene function and improved muscle function (Nelson *et al.*, 2016). More recent therapeutic approaches often take advantage of smaller Cas9 orthologs to facilitate single-AAV delivery (For the purpose of this thesis, single-AAV delivery and all-in-one AAV delivery are used interchangeably and refer to fitting Cas9, sgRNA and promoters into a single AAV). For example, single-AAV delivery of Cas9 has been used to target PCSK9 in the livers of adult mice to reduce serum cholesterol levels (Ran *et al.*, 2015; Ibraheim *et al.*, 2018b).

Despite this tremendous potential, there are disadvantages to AAV delivery. Its packaging limit constraints the size of the cargo so that some Cas9 orthologs (~ 1200 amino acids or larger) have difficulty being packaged along with their guides into a single vector. Furthermore, AAV persists for long periods in the target tissue, which may result in prolonged Cas9 expression and unintended off-targets. Finally, while immunogenicity is low, there is potential for the activation of the immune system at least by some serotypes. For example, some studies suggest that AAVs are detected by Toll-like receptors in mammals (Keeler and Flotte, 2019; Martino *et al.*, 2011; Zaiss *et al.*, 2002). Although some of these limitations apply regardless of the Cas9 ortholog, factors such as the size limit and increased off-target hinder some Cas9s more than others.

1.4.2 Currently-used Cas9 orthologs for AAV delivery

While at least a dozen Cas9 orthologs have been characterized, only a handful have been used for *in vivo* genome editing (Table 1.1). To date, in addition to SpyCas9 (type II-A), several <1,100 aa Cas9 orthologs have been validated for *in vivo* genome editing, including from strains of *N. meningitidis* (NmeCas9, 1,082 aa) (Ibraheim *et al*. 2018), *S. aureus* (SauCas9, 1,053 aa) (Ran *et al.,* 2015), *C. jejuni* (CjeCas9, 984 aa) (Kim *et al*., 2017), and *Streptococcus thermophilus (*Agudelo *et al.,* 2019*).* All existing approaches have limitations such as long PAMs and/or high off-targets, which will be discussed further in section 1.5 and addressed in chapters 3 and 4. The features of these Cas9s will be discussed below.

Species Name	PAM	Amenable to all-in-one AAV delivery?	Naturally accurate?
Streptococcus pyogenes (Spy)	NGG	✗	✗
Staphylococcus aureus (Sau)	NNGRRT	✓	✗
Campylobacter jejuni (Cje)	NNNNRYAC	✓	✓
Neisseria meningitidis (Nme)	NNNNGATT	✓	✓
Streptococcus thermophilus (st1)	NNAGAAW	✓	N/A

Table 1.1 Cas9 orthologs validated for genome editing by AAV
Five Cas9 orthologs have been validated for *in vivo* genome editing using AAV. All except for SpyCas9 are compact enough for single-AAV delivery, while only NmeCas9 and CjeCas9 are compact and less prone to off-target editing. However, the PAMs of the four compact orthologs are relatively long (underlined and shown in red boxes).

SpyCas9

Owing to its high efficiency and the fact that it was the first Cas9 characterized, SpyCas9 has been used extensively for AAV delivery. Despite its relatively large size and high number of off-targets, several different approaches have been used to deliver SpyCas9 via AAV. SpyCas9 and its sgRNA are ~ 4.2 kb in size, and when promoters and tags are added, it extends well above 5kb in length. Considering the packaging limit of ~4.7 kb for AAV and the reduced packaging efficiency of larger constructs (Keeler and Flotte, 2019), single-AAV delivery of SpyCas9 has been challenging. To overcome this, several approaches have been devised. For example, dual AAV delivery of Cas9 and sgRNA in separate vectors has been used to study gene function in the mouse brain (Swiech *et al.*, 2015). Such dual AAV delivery methods are more complex; both vectors must infect the same cell, and higher doses may be required to achieve similar outcomes. Furthermore, the current high cost of AAV production makes dual-AAV approaches less desirable. Finally, the long persistence of AAV (at least 6-12 months) results in prolonged expression of transgenes (Berns and Muzyczka, 2017) and in the case of SpyCas9, this results in higher propensity of off-target cleavage. Consequently, smaller and more accurate Cas9 orthologs have been characterized for AAV delivery.

NmeCas9

NmeCas9 has a canonical 5′-NNNNGATT-3′ PAM, though considerable variation from this consensus is permissive *in vitro* and several PAM variants are functional in mammalian cells as well (Amrani *et al.*, 2018b; Esvelt *et al.*, 2013; Hou *et al.*, 2013; Zhang *et al.*, 2015). In addition, NmeCas9 exhibits extremely clean off-target profiles in mammalian cells, even when editing sites whose guides are highly prone to off-targeting with wild-type SpyCas9 (Amrani *et al.*, 2018b; Lee, Cradick and Bao, 2016).More importantly, NmeCas9 single-AAV delivery has been used to target disease-causing loci *in vivo* (Ibraheim *et al.*, 2018b). Although NmeCas9's relatively long PAM limits the number of potential target sites in a locus of interest, its compact size and high accuracy make it an ideal candidate for AAV gene therapy applications.

CjeCas9

At 984 aa, CjeCas9 is among the smallest Cas9 orthologs identified to date. Its PAM was initially defined as NNNNACA (Fonfara *et al.*, 2014; Yamada *et al.*, 2017) and for several years, no reports emerged regarding its efficacy in mammalian genome editing. More recently, Kim *et al*. reassessed the PAM requirements and found an additional nucleotide (NNNNRYAC) to contribute to editing efficiency (Kim *et al.*, 2017). When genome-wide CjeCas9 editing specificity was compared to that of SpyCas9 and SauCas9 (using sites that were edited at comparable frequencies), CjeCas9 showed much lower off-target cleavage at multiple tested sites (on average 10-fold less). This report further established CjeCas9 all-in-one AAV vectors as effective

delivery vehicles for genome editing, e.g. in the eye, again with little or no off-targeting, even after eight months post-delivery (Kim *et al.*, 2017). More recently, CjeCas9's *in vivo* AAV application was expanded to rescue dystrophin deficiency in a mouse model of DMD (Koo *et al.*, 2018). The small size of CjeCas9 allows for the packaging of the Cas9, sgRNA and DNA donors that are several hundred nucleotides in length and has the potential to be used for single-AAV HDR applications.

SauCas9

SauCas9 was the first compact Cas9 ortholog developed for single-AAV genome editing and has been widely used (Ran *et al.*, 2015). It has an NNGRRT PAM and at 1053 amino acids, it is the amenable to single-AAV delivery. SauCas9 falls in the middle of the accuracy spectrum; while it is not as accurate as some type II-C orthologs such as NmeCas9 and CjeCas9, it has fewer off-targets than its type II-A counterpart SpyCas9. These properties have made it the second most widely used ortholog; SauCas9 single-AAV vectors have been used in mouse models of diseases such as DMD and liver metabolic diseases (Nelson *et al.*, 2016; De Caneva *et al.*, 2019). However, the complex PAM of SauCas9 limits its applications. A SauCas9 mutant ($SauCas9^{KKH}$) that has reduced PAM constraints (N_3RRT) has been developed, though this increase in targeting range often comes at the cost of reduced on-target editing efficacy, and off-target edits are still observed (Kleinstiver *et al.*, 2015).

St1Cas9

St1Cas9 is the latest addition to the single-AAV CRISPR toolbox. St1Cas9 is 1121 amino acids in length and recognizes NNAGAA and NNGGAA PAMs in mammalian cells. Single-AAV Delivery of St1Cas9 to the mouse liver led to the phenotypic rescue of a mouse model of hereditary tyrosinemia. However, the PAM is relatively long, similar to the other compact orthologs described above. Preliminary studies suggest that it may be more accurate than SpyCas9 (Muller *et al.*, 2016), the genome-wide off-target profile of St1Cas9 remains to be investigated.

Taken together, while all the Cas9 orthologs mentioned above have their advantages, none satisfies all the requirements to be the ideal ortholog for AAV delivery. These limitations will be further discussed below and addressed in chapters 3 and 4 of this thesis.

1.5 Current limitations of CRISPR-Cas9

The safety and efficacy of CRISPR-Cas9 is of paramount importance before its widespread use in therapeutics. There are limitations that apply to almost all Cas9 orthologs. The bacterial origin of Cas9 and its relatively large size mean that the human body could mount an immune response against it. Furthermore, many of the characterized Cas9s to date are from human commensal/pathogenic bacteria, with the possibility of pre-existing antibodies against some of these orthologs. Indeed, analysis of 125 blood donors revealed that more than half of the donors had antibodies against two of the most commonly used Cas9s, namely SauCas9 and SpyCas9 (Charlesworth *et al.*, 2019). Although this is probably not a concern for orthologs that originate from other sources (e.g. GeoCas9 (Harrington *et al.*, 2017)), the immune response to Cas9 must be studied and remedied before *in vivo* clinical use.

Another major hurdle has been the number of off-targets, i.e. Cas9 inducing DSBs at sites similar to the intended target, which may have undesired deleterious consequences. Wild-type SpyCas9, the most widely used for editing, has a high degree of off-target activity in most cell types. There has been extensive characterization and engineering to minimize the tendency to edit such near-cognate, off-target sites (Tsai and Joung, 2016; Tycko, Myer and Hsu, 2016; Chen *et al.*, 2017). Such engineered Cas9s and more naturally accurate orthologs promise to pave the way for the future of genome editing. Nevertheless, regardless of the Cas9 ortholog used, careful selection of

target sites and extensive evaluation of off-targeting is essential for safe genome editing.

In addition to off-targets, there is an inherent risk associated with creating DSBs in the genome. Repair outcomes are often unpredictable and context dependent. Specifically, in line with the dependence of Cas9 on cellular repair machinery, certain cell types are inherently recalcitrant to genome editing, particularly to HDR (Jasin and Haber, 2016). For example, precise editing of post-mitotic neurons has been challenging, and while AAV-mediated HDR has been achieved (Nishiyama, Mikuni and Yasuda, 2017), the efficiency varies in different brain regions. In some contexts, such as in human pluripotent stem cells, Cas9-induced DSBs activate the P53 repair pathway with detrimental consequences, including cell death (Haapaniemi *et al.*, 2018).

Finally, delivery of CRISPR components to the target tissue of interest remains a challenge. While the delivery problem is not unique to Cas9, the relatively large size of some Cas9s makes delivery even more cumbersome. Several delivery avenues including viral and non-viral vectors have been tested to deliver Cas9 in mice. Nevertheless, delivering Cas9 to the desired tissue remains a challenge.

In summary, while CRISPR-Cas9 has the potential to improve our understanding of biology and treat diseases, *in vivo* applications are limited due to potential safety issues and pitfalls of current tools. In this thesis, we will present our efforts to overcome some of these hurdles by identifying new Cas9 orthologs and their off-switches.

CHAPTER 2: Phage-encoded inhibitors of CRISPR-Cas9

2.1 Introduction:

As discussed previously, CRISPR is an RNA-guided system, where Cas proteins are programmed to target and destroy phage nucleic acids. CRISPR-Cas adaptive immunity is present in 90% of archaea and ~50% of all bacteria. As part of the constant arms race between bacteria and phages, that phages have evolved anti-CRISPR strategies to overcome CRISPR immunity. While phages can temporarily escape the sequence-specific targeting by CRISPR through mutations, evolutionary studies showed that the only way for a phage to escape extinction is by the expression of an anti-CRISPR gene (van Houte *et al.*, 2016). In other words, mutations alone are likely not enough to overcome CRISPR immunity. Anti-CRISPR (Acr) proteins provide one way to disarm CRISPR-Cas systems by directly inhibiting interference. The first Acrs were discovered in phages that successfully infect *Pseudomonas aeruginosa* strains despite the presence of an active type I CRISPR-Cas system (Bondy-Denomy *et al.*, 2013). While Acrs themselves are often poorly conserved, they are frequently encoded upstream of transcriptional regulator genes known as anti-CRISPR-associated (aca) genes (Pawluk *et al.*, 2014).

In addition to the biologic significance of discovering Acrs against Cas9, such Acrs may have the potential to modulate Cas9's activity in genome engineering applications. For example, off-target cleavage resulting from prolonged Cas9 expression can be remedied by expressing an Acr off-switch to limit the duration of

editing. HDR rates can be enhanced by limiting Cas9's activity for precise editing, by inhibiting Cas9 during the G1 phase of the cell cycle, when HDR pathways are suppressed in most cell types (Orthwein *et al.*, 2015). Finally, Cas9 has been used for gene drives, i.e., propagation of a desired gene via non-Mendelian forced inheritance (Gantz *et al.*, 2015). However, it is imperative to be able to switch off Cas9 if needed, and Acrs can be used for this purpose. This chapter will describe the discovery of the first type II-specific Acrs in 2016 (Pawluk *et al.*, 2016) and our efforts since to uncover their mechanistic and evolutionary diversity.

2.2 Results:

2.2.1 The discovery of Anti-CRISPR proteins against NmeCas9

While the first Acr proteins were reported in 2013, none had been reported to inhibit type II CRISPR-cas systems that used Cas9 as their effector enzyme (Pawluk, Davidson and Maxwell, 2018). This is in part due to the lack of Acr conservation. However, the lab of Dr. Alan Davidson at the University of Toronto took advantage of the conservation of *aca* genes and found that they are conserved among diverse phyla, some of which possessed Cas9s. Using previously discovered Aca1 and Aca2 (from type I Acrs), the Davidson lab identified a candidate anti-CRISPR gene in a *Brackiella oedipodis* putative conjugative element that encoded a small hypothetical protein which was directly upstream of an *aca2* gene (Figure 2.1A). This putative anti-CRISPR had several homologs encoded in MGEs of diverse bacteria, including *Neisseria meningitidis.* The aca associated with this Acr (*aca3*) was used to find two other Acr proteins in other meningococcal strains, bringing the total to three putative Acrs. Because *N. meningitidis* has the best-established type II-C CRISPR-Cas system and our lab had worked on it for several years, it was the ideal system to test these proteins' ability to inhibit Cas9.

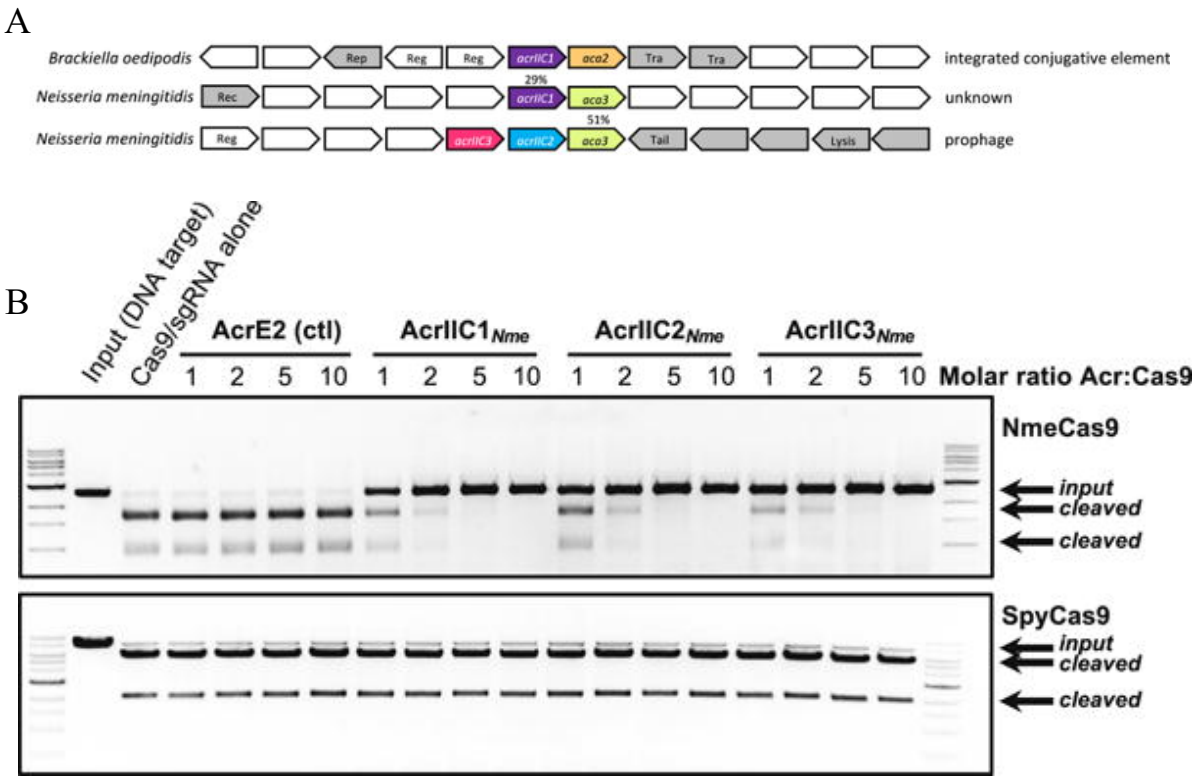


Figure 2.1: Putative type II-C acrs specifically inhibit NmeCas9 *in vitro*
(A) Schematic representation of candidate type II-C acr and aca genes within putative MGEs in the genomes of strains of B. oedipodis and N. meningitidis. Homologous genes are color-matched, with percent amino acid identities indicated. Annotation of relevant genes is as follows: Rep, plasmid replication protein; Reg, transcriptional regulator; Tra, conjugal transfer protein; Rec, recombinase; Tail, phage tail structural protein; Lysis, phage lysis cassette. Genes colored in gray have MGE-related functions and/or show clear evidence of horizontal transfer.
(B) Linearized plasmid DNA bearing a protospacer adjacent to a PAM sequence was subjected to *in vitro* digestion by purified, recombinant, sgRNA-programmed NmeCas9 (upper panel) or SpyCas9 (lower panel). Where indicated at the top of each lane, Cas9 was pre-incubated with purified anti-CRISPR proteins as indicated with AcrE2 as a negative control. Molar equivalents of anti-CRISPR protein (relative to Cas9) are shown at the top of each lane, and mobility of input and cleaved DNAs are denoted on the right. The NmeCas9 cleavage assays shown are representative of three independent replicates.

To address this question, we tested the ability of Cas9 to cleave targets *in vitro* using recombinant NmeCas9 in the presence of these Acrs. The reaction consists of forming a complex between an *in vitro* transcribed sgRNA and recombinant Cas9, followed by the addition of each Acr protein at different molar ratios, and finally adding the target DNA. We used a type I-E inhibitor (AcrE2) (Pawluk *et al.*, 2014) as a negative control Acr (Figure 2.1B) All three families of Acr proteins inhibited NmeCas9 in a dose-dependent manner (Figure 2.1B, top). We also found that these Acrs are highly specific, since they were unable to inhibit SpyCas9 (~25% homology) at even 10:1 molar ratio (Figure 2.1B, bottom). These results signify the discovery of the first three families of Acrs against Cas9, named AcrIIC1, AcrIIC2 and AcrIIC3.

2.2.2 Anti-CRISPRs are found in diverse bacterial species:

The initial discovery of Acrs against Cas9 paved the way for the discovery of such proteins in other bacterial strains. Again, work spearheaded by the Davidson lab led to the discovery of Acrs associated with *aca2* in two species: *Haemophilus parainfluenzae* and *Simonsiella muelleri.* These aca orthologs share 38% and 36% identity to *B. oedipodis* aca2, respectively, and 51% identity to each other (Figure 2.2A), with putative Acrs upstream. Both strains encode predicted type II-C CRISPR-Cas machineries with Cas9 orthologs that exhibit 59% and 62% identity with NmeCas9, respectively. Based on these similarities and our previous work indicating that type II anti-CRISPRs could inhibit Cas9 orthologs outside their host species/strain, we first tested for Acr activity against the well-characterized NmeCas9 *in vitro*. We cloned each candidate Acr sequence into a bacterial expression vector, purified recombinant proteins from *E. coli*, and tested their abilities to prevent DNA cleavage by NmeCas9 *in vitro* at a 10:1 ratio of Acr:Cas9-sgRNA. We found that both Acrs efficiently inhibited NmeCas9, corroborating the existence of Acr proteins in diverse phages. As per the unified resource for naming Acr proteins (Bondy-Denomy *et al.*, 2018), we named these AcrIIC4$_{Hpa}$ and AcrIIC5$_{Smu}$.

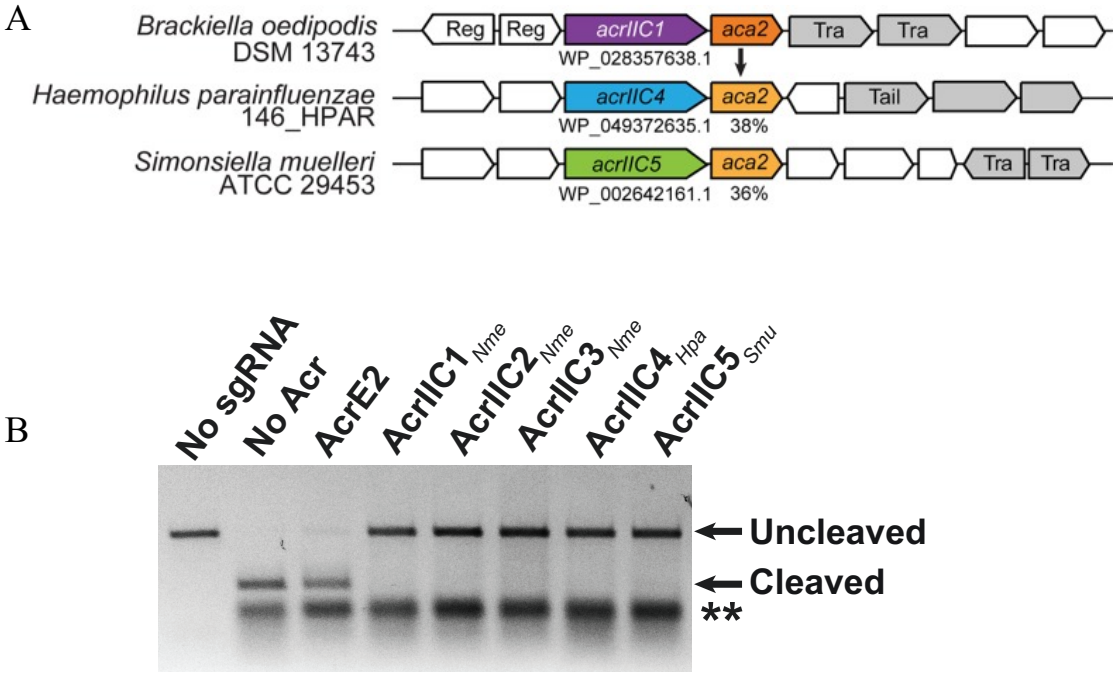


Figure 2.2 Identification and *in vitro* validation of two anti-CRISPR protein families.

(A) Schematic of candidate anti-CRISPR proteins and *aca2* genes in the genomic context of *H. parainfluenzae* (AcrIIC4$_{Hpa}$) and *S. muelleri* (AcrIIC5$_{Smu}$). Gray genes are associated with mobile DNA, and known gene functions are annotated as follows: "Reg" is a transcriptional regulator, "Tail" is involved in phage tail morphogenesis, and "Tra" is a transposase. The B. oedipodis *aca2* gene is used as a query for pBLAST searches, and percent identities of *aca2* orthologs are denoted. Arrows are not drawn to scale.

(B) *In vitro* cleavage of target DNA by the NmeCas9-sgRNA complex in the presence of anti-CRISPR protein. Preformed NmeCas9-sgRNA RNP complex was incubated with purified anti-CRISPR proteins as indicated with AcrE2 as a negative control, AcrIIC1 as a positive control, and candidate Acrs. Then, a linearized plasmid with a protospacer and PAM sequence was added to the reaction mixture. ** depicts the sgRNA band.

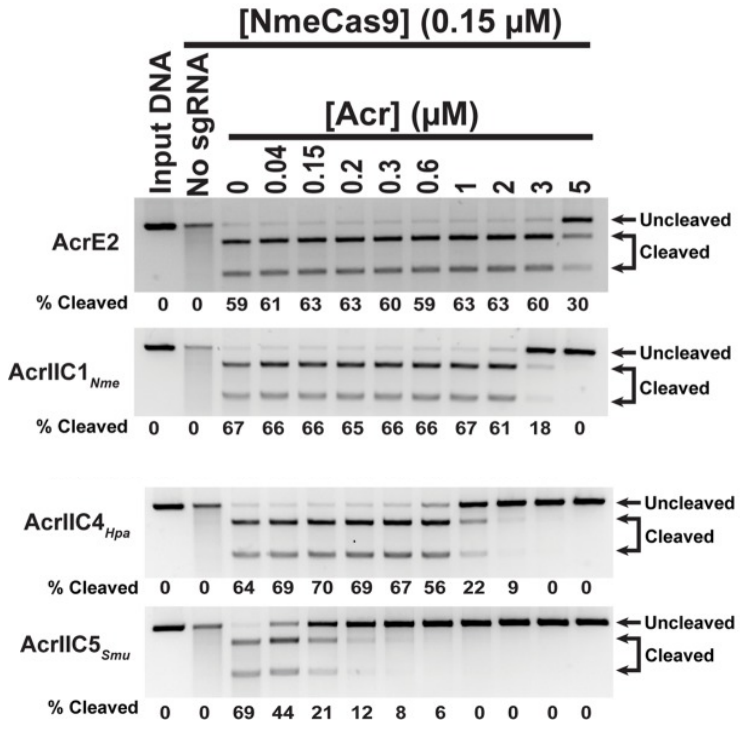
To compare the potency of these Acrs with the previously discovered AcrIIC1 family, we performed a similar *in vitro* cleavage assay with increasing concentrations of Acrs. When each of the purified candidate Acrs was added to parallel reaction mixtures, cleavage was inhibited in a concentration-dependent manner, with complete inhibition being reached at ~20-fold ($AcrIIC4_{Hpa}$) and ~7-fold ($AcrIIC5_{Smu)}$ molar excess of Acrs (Figure 2.3). Incubation with our negative control AcrE2 did not affect target DNA cleavage by NmeCas9.

Once we confirmed the anti-CRISPR inhibition of sgRNA-guided NmeCas9 DNA cleavage *in vitro*, we then addressed the mechanisms of NmeCas9 inhibition by $AcrIIC4_{Hpa}$ and $AcrIIC5_{Smu}$. We tested whether the new Acrs prevent sgRNA loading or DNA target binding. First, we checked whether sgRNA loading onto NmeCas9 is inhibited by either anti-CRISPR. We carried out electrophoretic mobility shift assays (EMSAs) by incubating NmeCas9 and sgRNA with or without Acr, and then visualizing sgRNA mobility after native gel electrophoresis by SYBR Gold staining. In the absence of any anti-CRISPR, incubation of NmeCas9 with its cognate sgRNA resulted in a gel shift that indicates formation of a stable RNP complex (Fig 2.3B) When NmeCas9 was incubated with a negative-control anti-CRISPR (AcrE2) before the addition of sgRNA, NmeCas9:sgRNA complex formation was unaffected. Similarly, when incubated with $AcrIIC4_{Hpa}$ and $AcrIIC5_{Smu}$, efficient NmeCas9:sgRNA complex formation was again observed, suggesting that neither Acr protein significantly affected RNP assembly.

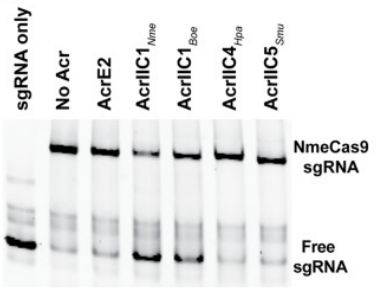
To test if target DNA engagement by the NmeCas9:sgRNA complex is prevented by either AcrIIC4$_{Hpa}$ or AcrIIC5$_{Smu}$, we performed EMSAs and fluorescence polarization assays after incubating the RNP with each Acr, before adding target DNA To inhibit DNA target cleavage, we omitted divalent metal ions from the reaction mixtures. The target DNA exhibited the expected mobility shift in the absence of Acr, or in the presence of AcrE2 or AcrIIC1$_{Nme}$ (as controls). Our results indicate that both AcrIIC4$_{Hpa}$ and AcrIIC5$_{Smu}$ prevented NmeCas9 RNP binding to the target DNA.

We then tested the ability of these Acrs to inhibit NmeCas9 genome editing in mammalian cells. To do this, we cotransfected HEK293T cells transiently with plasmids expressing anti-CRISPR protein, NmeCas9 and sgRNAs targeting genomic sites. We then used targeted deep sequencing to estimate genome editing efficiency (Kai Xiong et al, 2013). In agreement with our *in vitro* data, expression of all five ACR families inhibited NmeCas9-mediated mutagenesis to undetectable levels. In contrast, they had no effect on SpyCas9 genome editing at the same site. Indel frequencies were measured by targeted deep sequencing of PCR-amplified genomic DNA collected after transfection of NmeCas9 plasmid with or without Acrs at different dosages.

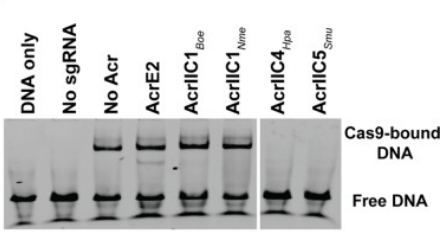
A


[NmeCas9] (0.15 μM)
Input DNA
No sgRNA
[Acr] (μM)
0
0.04
0.15
0.2
0.3
0.6
1
2
3
5
AcrE2
Uncleaved
Cleaved
% Cleaved 0 0 59 61 63 63 60 59 63 63 60 30
AcrIIC1Nme
Uncleaved
Cleaved
% Cleaved 0 0 67 66 66 65 66 66 67 61 18 0
AcrIIC4Hpa
Uncleaved
Cleaved
% Cleaved 0 0 64 69 70 69 67 56 22 9 0 0
AcrIIC5Smu
Uncleaved
Cleaved
% Cleaved 0 0 69 44 21 12 8 6 0 0 0 0

B


sgRNA only
No Acr
AcrE2
AcrIIC1Nme
AcrIIC1Boe
AcrIIC4Hpa
AcrIIC5Smu
NmeCas9 sgRNA
Free sgRNA

C


DNA only
No sgRNA
No Acr
AcrE2
AcrIIC1Boe
AcrIIC1Nme
AcrIIC4Hpa
AcrIIC5Smu
Cas9-bound DNA
Free DNA

D

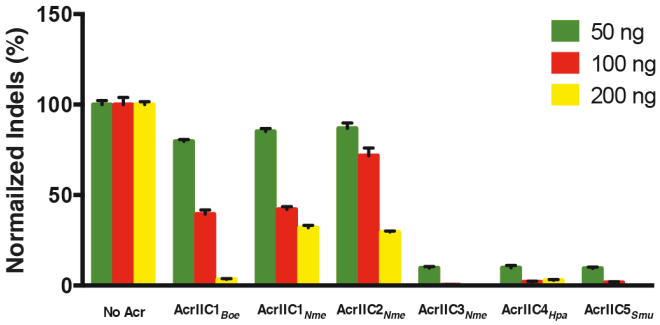

Normalized Indels (%)
150
100
50
0
50 ng
100 ng
200 ng
No Acr
AcrIIC1Boe
AcrIIC1Nme
AcrIIC2Nme
AcrIIC3Nme
AcrIIC4Hpa
AcrIIC5Smu

Figure 2.3 The new Acr families inhibit NmeCas9 *in vitro* and in cells
(A) *In vitro* cleavage of target DNA by the NmeCas9-sgRNA complex in the presence of anti-CRISPR protein.
Preformed NmeCas9-sgRNA RNP complex was incubated with purified anti-CRISPR proteins as indicated with AcrE2 as a negative control, AcrIIC1 as a positive control, and candidate Acrs. Then, a linearized plasmid with a protospacer and PAM sequence was added to the reaction mixture. Molarities of anti-CRISPR protein (relative to constant Cas9 molarity) are shown at the top of each lane, mobilities of input and cleaved DNAs are denoted on the right, and cleavage efficiencies ("% cleaved") are given at the bottom of each lane.
(B and C) A native gel of the sgRNA visualized by SYBR gold staining (B) and of the FAM-labeled target DNA (C), both of which were added last to NmeCas9 + Acr (and in panel C, + sgRNA) incubation.
(D) Indel frequencies measured by targeted deep sequencing of PCR-amplified genomic DNA collected after transfection of NmeCas9 plasmid with or without Acrs at different dosages.

2.2.3 AcrIIC1 is a broad-spectrum inhibitor of CRISPR Cas9

AcrIIC1 putative orthologs identified by PSI-BLAST suggest that it may be part of a family of proteins in diverse bacterial species (Pawluk *et al.*, 2016). This diversity is mirrored in the Cas9 orthologs from the bacterial genomes containing AcrIIC1 family of Acrs in prophages. We hypothesized that AcrIIC1 may be promiscuous with respect to the Cas9 orthologs it can inhibit. We had already established that it was unable to inhibit SpyCas9 (type II-A), so we focused on the two other type II-C orthologs that had been validated for genome editing, namely CjeCas9 and GeoCas9, which are 36% and 42% identical to NmeCas9 respectively (Kim *et al.*, 2017; Harrington *et al.*, 2017). *In vitro* cleavage data shows that in addition to NmeCas9, the AcrIIC1 from *N. meningitidis* (AcrIIC1Nme) exhibits robust inhibition of GeoCas9 and CjeCas9 (figure 2.4B). By contrast, AcrIIC2Nme and AcrIIC3Nme were both highly specific for NmeCas9, having no noticeable impact on CjeCas9- or GeoCas9-catalyzed DNA cleavage.

To determine whether inhibition by AcrIIC1 can disable CjeCas9 in genome-editing applications, we transfected HEK293T cells with plasmids expressing NmeCas9, CjeCas9 or SpyCas9 and their respective sgRNAs targeting identical loci in the presence or absence of AcrIIC1 (figure 2.4C). Similar to the *in vitro* cleavage assays, CjeCas9 was inhibited by AcrIIC1 but not by AcrIIC3. Expressing $AcrIIC1_{Nme}$ or the $AcrIIC1_{Boe}$ (not shown) resulted in efficient inhibition of CjeCas9, indicating that this promiscuity is not unique to the $AcrIIC1_{Nme}$ ortholog. In a similar experiment, the Doudna lab showed that AcrIIC1 is also a potent inhibitor of GeoCas9 in mammalian

cells. The robust inhibition of both CjeCas9 and GeoCas9, in addition to NmeCas9, suggested that AcrIIC1 exploits a conserved feature of the Cas9 protein.

Mechanistic and structural studies by the Doudna lab show that AcrIIC1 family indeed binds to a conserved region of Cas9's active site. A 1.5-Å resolution crystal structure of $AcrIIC1_{Nme}$ bound to the HNH domain of NmeCas9 revealed that AcrIIC1 binds directly to the HNH active site (Figure 2.4E), restricting it from accessing the target DNA. AcrIIC1 binds to the active site interface of the HNH domain through several ionic and hydrogen-bonding interactions. Mapping amino acid conservation onto the structure revealed that residues within the binding interface of both the HNH domain and AcrIIC1 are highly conserved (Figures 2.4D). In contrast to this observed conservation, antagonistic binding interfaces often evolve rapidly, leading to lower conservation (Franzosa and Xia, 2011), suggesting that AcrIIC1 is targeting a highly conserved surface to limit the chance for the host to escape inhibition.

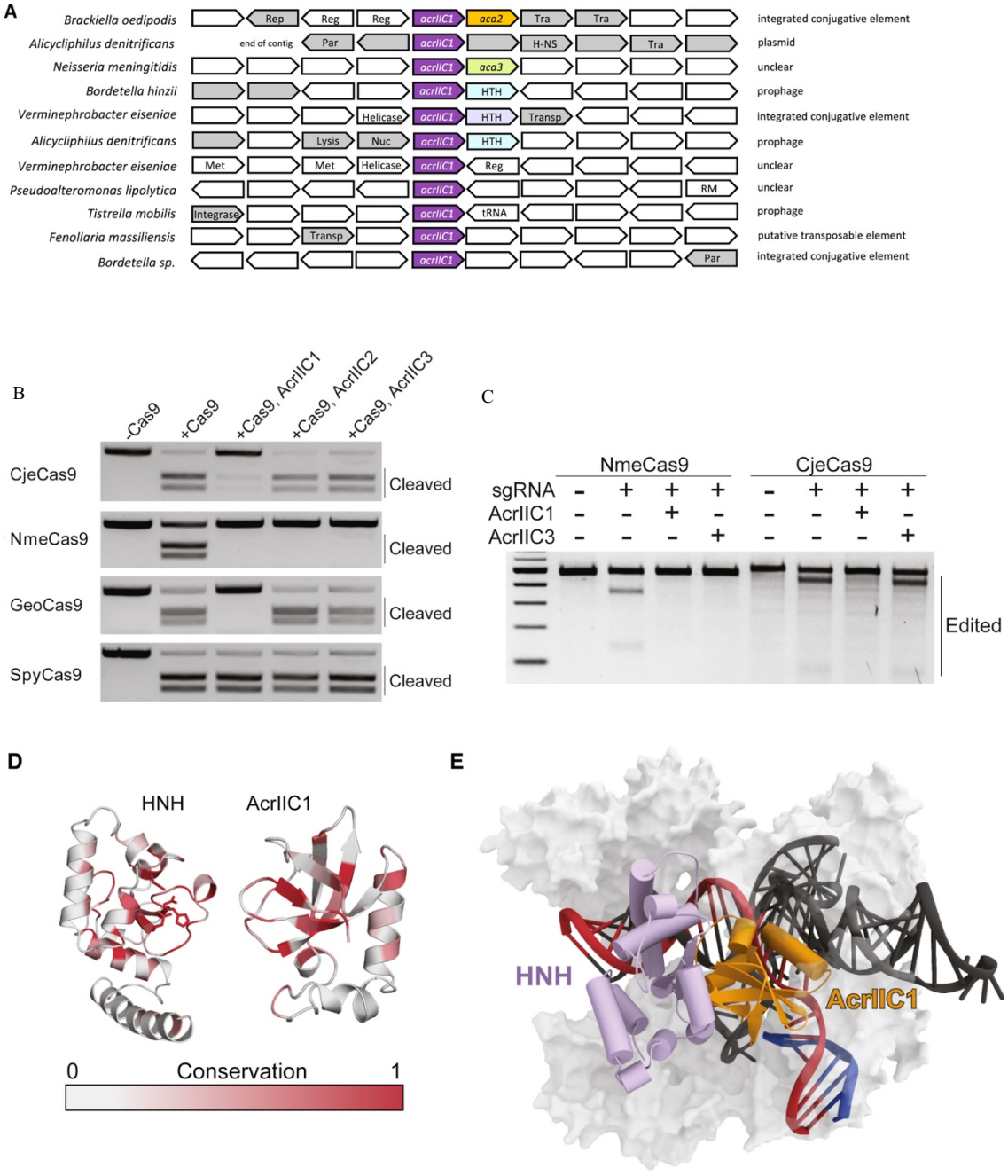

A
Brackiella oedipodis
Alicycliphilus denitrificans
Neisseria meningitidis
Bordetella hinzii
Verminephrobacter eiseniae
Alicycliphilus denitrificans
Verminephrobacter eiseniae
Pseudoalteromonas lipolytica
Tistrella mobilis
Fenollaria massiliensis
Bordetella sp.
Rep
Reg
Reg
acrIIC1
aca2
Tra
Tra
integrated conjugative element
end of contig
Par
H-NS
Tra
plasmid
aca3
unclear
HTH
prophage
Helicase
HTH
Transp
integrated conjugative element
Lysis
Nuc
HTH
prophage
Met
Met
Helicase
Reg
unclear
RM
unclear
Integrase
tRNA
prophage
Transp
putative transposable element
Par
integrated conjugative element
B
-Cas9
+Cas9
+Cas9, AcrIIC1
+Cas9, AcrIIC2
+Cas9, AcrIIC3
CjeCas9
NmeCas9
GeoCas9
SpyCas9
Cleaved
C
NmeCas9
CjeCas9
sgRNA
AcrIIC1
AcrIIC3
Edited
D
HNH
AcrIIC1
0
Conservation
1
E
HNH
AcrIIC1

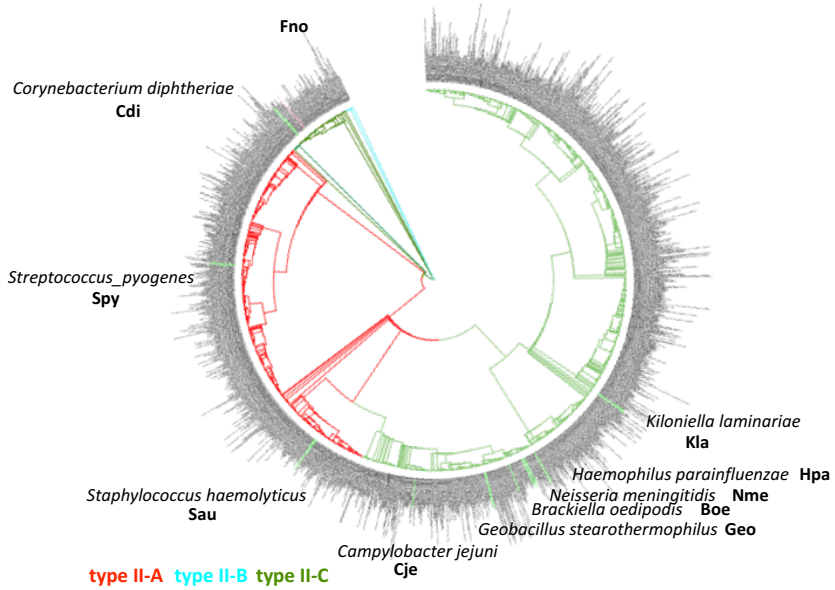
Figure 2.4. AcrIIC1 inhibits Cas9 orthologs by binding to the HNH domain
(A)Schematic representation of putative AcrIIC1 orthologs identified by PSI-BLAST and their genomic contexts. The species in which each is found and its predicted genomic region classification (i.e. prophage, integrated conjugative element) are indicated. Gene arrows are not drawn to scale.
(B) DNA cleavage assays conducted by several Cas9 orthologs in the presence of AcrIIC1, AcrIIC2, and AcrIIC3 (−Cas9, no Cas9 added; +Cas9, Cas9 and sgRNA added.
(C) T7E1 assay analyzing indels produced by CjeCas9 and NmeCas9 shows that CjeCas9 genome editing is inhibited by AcrIIC1Nme, but not AcrIIC3Nm in HEK293T cells.
(D) Binding interfaces of NmeCas9 HNH domain and AcrIIC1 show residue conservation. Conservation was calculated using multiple sequence alignments of AcrIIC1 orthologs and Cas9 HNH domains. Conserved residues are colored red (1, 100% sequence identity), and non-conserved residues are colored white (0). (E) Model of AcrIIC1 inhibiting cleavage of both target and non-target strands. NmeCas9 HNH domain (purple) was modeled into a "docked" position using dsDNA-bound SpyCas9 structure (PDB: 5F9R) as a reference for a homology model of NmeCas9. Placement of AcrIIC1 (orange) between the HNH domain and the target strand (red) prevents target cleavage and activation of the RuvC domain for non-target strand (dark blue) cleavage.

2.2.4 Acrs can be highly specific or broad-spectrum

Our results on AcrIIC1 prompted us to test all previously reported type II-C Acrs against a diverse set of Cas9 orthologs. In addition to CjeCas9 and GeoCas9, we tested the Acrs on Cas9s from *Haemophilus parainfluenzae* (HpaCas9), *Brackiella oedipodis* (BoeCas9) and *Kiloniella laminariae* (KlaCas9) with 66%, 46% and 24% sequence homology to NmeCas9 (See chapter 3 for in-depth characterization of these orthologs). First, we performed *in vitro* cleavage of a target DNA using CjeCas9 in the presence of previously described Acr proteins. Based on the previous data suggesting that CjeCas9 is only inhibited by AcrIIC1 family and not by AcrIIC2 or AcrIIC3, we focused only on AcrIIC4 and AcrIIC5, and used AcrIIC1 as a positive control. Our *in vitro* results corroborated that AcrIIC1 inhibits CjeCas9, while neither AcrIIC4 nor AcrIIC5 prevented target DNA cleavage by CjeCas9. We also performed the same *in vitro* cleavage assay on all the Cas9s mentioned above, using AcrIIC1 as a control. A protospacer was cleaved *in vitro* using the Cas9-sgRNA complex of each CRISPR system in the presence or absence of Acrs, and the product was run on an agarose gel to visualize cleaved vs. uncleaved bands similar to figure 2.2B. The results are summarized in table 2.1.

While AcrIIC1 family inhibited all type II-C orthologs tested, the other four families of Acrs are relatively specific. All five families inhibited NmeCas9, HpaCas9 and SmuCas9, suggesting that the target potential binding site(s) for each Cas9 is conserved among these three orthologs. For example, it was recently shown that AcrIIC2 inhibits sgRNA loading by binding to the bridge helix of NmeCas9, which is

not highly conserved among type II-C orthologs (Thavalingam *et al.*, 2019). In addition to the three closely-related orthologs, AcrIIC4 is also able to inhibit BoeCas9, suggesting a more diverged area being targeted. Taken together, these data suggest that Acrs can be broad-spectrum or specific and pave the way to detailed mechanistic studies.



Type II-C Acr family

Cas9	Identity to NmeCas9 (%)	IIC1	IIC2	IIC3	IIC4	IIC5
NmeCas9	100					
HpaCas9	66					
SmuCas9	62					
BoeCas9	46					
GeoCas9	42					
CjeCas9	36					
KlaCas9	28					

Figure 2.5. *in vitro* cleavage with diverse Cas9s in the presence of AcrIIC families
(A) Phylogenetic tree of Cas9, those with more than 90% identity have been depicted as one. The Cas9s used previously and in this study are marked. Clades are colored by Cas9 sub-type.
(B) Summary of *in vitro* cleavage data by each Cas9 in the presence of type II-C Acr families. Blue indicates inhibition, yellow indicates lack of inhibition.

2.3 Discussion

We have shown widespread occurrence of type II-C Acrs across diverse bacterial phyla. While the evolutionary origin of Acrs remain unclear, their widespread presence suggests a strong selective pressure imposed by CRISPR-Cas9. It is also unclear how the long-term exposure to Acrs shaped the evolution of CRISPR. There is undoubtedly a staggering diversity among Acr proteins, presumably even more than their CRISPR counterparts. The pace of phage evolution and the lack of conservation among Acr proteins suggest a rapid, convergent evolution against CRISPR.

Our guilt-by-association approach is limited due to its dependence on the presence of the *aca* gene. To address this bias, other methods of Acr discovery have been implemented by groups throughout the world. For example, an approach was developed to find bacterial genomes with CRISPR systems that contained self-targeting spacers, i.e. spacers that should be lethal due to targeting the bacterial genome (Rauch *et al.*, 2017). The presence of such spacers indicates a mechanism, possibly an Acr, by which the CRISPR system is deactivated. This approach led to the discovery of several Acr proteins in *Listeria monocytogenes* that target type II-A CRISPR-Cas9 systems. Using these various approaches, there are about 40 different families of Acrs discovered to date (Stanley and Maxwell, 2018). Nevertheless, considering the perpetual fight between phages and bacteria/archaea, we have definitely only scratched the surface of Acr diversity. Future Acr studies will not only focus on the diversity of these proteins, but the outcome of a CRISPR system once it has been inhibited by an Acr. Several intriguing possibilities exist regarding the bacterial response: 1) The

bacterium could respond by evolving the CRISPR component that is the target of the Acr on an evolutionary timescale. For example, if the Acr binds to the PID of Cas9, there could be selective pressure to evolve a novel PID. However, in cases where the Acr binds to a highly conserved region (such as AcrIIC1 binding to the HNH catalytic residues), evolutionary escape may not be an easy option. 2) The bacterium could respond by evolving proteins to disarm the Acr, an anti-anti-CRISPR. However, no such enzyme has been reported to date. 3) Finally, the presence of an Acr may render CRISPR obsolete and result in the elimination of selective pressure to maintain a CRISPR system. In fact, as much as 12% of CRISPR systems are incomplete or mutated (Makarova *et al.*, 2015), which may be in part due to the presence of Acrs . For example, the CRISPR locus of *Simonsiella muelleri*, which co-exists in the same genome that AcrIIC5$_{Smu}$ is derived from, appears to be degenerate (see chapter 3 for details).

Acrs exhibit broad mechanistic diversity. In addition to binding to the HNH domain like AcrIIC1, some of the strategies include mimicking DNA, preventing sgRNA loading, and inhibition of target cleavage (Stanley and Maxwell, 2018). Interestingly, no Acr has been found to inhibit adaptation, the first stage of CRISPR immunity. This may be due to the fact that by targeting interference, Acrs ensure that any RNP complex already present from previous infections gets inhibited. Furthermore, at least in type II-A systems, Cas9 is essential for adaptation and it is very likely that by inhibiting Cas9 (Heler *et al.*, 2015), both adaptation and interference are inactivated, and not interference alone.

In addition to the biological significance of Acrs against Cas9, their potential applications in biotechnology are just being realized. For example, while Cas9-facilitated gene drive is a powerful means of potentially controlling populations of insect-borne diseases such as malaria, the possibility of runaway drive has hindered current applications and tools such as Acrs may prove beneficial in controlling drive. In 2018, Basgall et al. reported the use of Acrs in inhibiting gene drives in yeast (Basgall *et al.*, 2018). Another application of Acrs was recently described by our group, where Acrs restricted genome editing *in vivo* to a specific tissue by silencing Cas9 in ancillary tissues. Finally, Acrs that inhibit DNA binding of Cas9 can be used for dCas9-based tools. A recent study took advantage of a type IIA Acr to temporally restrict dCas9-Tet1 activity to modulate the timing of methylation (Liu *et al.*, 2018). Overall, we have only begun to appreciate the diversity and complexity of Acrs.

2.4 Materials and methods

Vectors used in this chapter

NmeCas9 was cloned into the pMCSG7 vector containing a T7 promoter followed by a 6XHis tag and a tobacco etch virus (TEV) protease cleavage site using Gibson Assembly cloning. The GeoCas9-expressing plasmid (expressing the GeoCas9 ortholog from *G. stearothermophilus* strain ATCC 7953) was obtained from Addgene (#87700) and similarly cloned into the pMCSG7 vector. For construction of sgRNA scaffolds for GeoCas9, the tracrRNA was predicted by crRNA repeat complementarity as well as homology to the NmeCas9 tracrRNA. These sgRNA scaffolds were ordered as gBlocks (IDT) along with overhangs to clone into pLKO.1 plasmid using Gibson Assembly. The CjeCas9 sgRNA was also cloned into pLKO based on the sequence from Kim et al. All sgRNA scaffolds were used as the templates to create *in vitro*-transcribed sgRNAs.

DNA sequences encoding candidate anti-CRISPR proteins were synthesized and cloned into a pUC57 mini (AmpR) vector with an *N. meningitidis* 8013 Cas9 promoter sequence for bacterial work. For anti-CRISPR protein purification, the Acr insert was amplified and inserted into the pMCSG7 backbone by Gibson Assembly (NEB), resulting in pMCSG7-Acr.

For editing of genomic dual target sites by both SpyCas9 and NmeCas9, we used Cas9 and cognate sgRNA expression vectors that were described previously in Pawluk et al. (2016). To generate the Acr expression vector, the Acr ORF was

amplified from pUC57-Acr and inserted into XhoI-digested pCSDest2 by Gibson Assembly.

Expression and purification of NmeCas9

The NmeCas9 pMCSG7 construct (deposited on Addgene #71474) was transformed into the Rosetta 2 DE3 strain of *E. coli.* Expression of NmeCas9 was performed as previously described for SpyCas9 with some modifications (Jinek 2012). Cells were grown in Terrific Broth (TB) medium at 37C to an optical density (OD600 nm) of 0.6-0.8 Protein expression was induced by the addition of 1 mM IPTG for 16 hr at 16C. Cells were lysed by sonication in 50 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.5 mM DTT and 5% glycerol supplemented with 0.5 mM PMSF, lysozyme and protease inhibitor cocktail (Sigma). Clarified lysates were bound in batch to Ni-NTA agarose (QIAGEN), and bound protein was eluted with 300 mM imidazole. Purified Cas9:sgRNA was dialyzed into 20 mM HEPES pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM DTT and 1 mM PMSF) for protein interaction experiments.

Expression and purification of Acr proteins

6×His-tagged Acrs were expressed in *E. coli* strain BL21 Rosetta (DE3). Cells were grown in LB at 37°C to an optical density (OD_{600}) of 0.6 in a shaking incubator. Next, the bacterial cultures were immediately cooled to 18°C, and protein expression was induced by adding 1 mM IPTG. Bacterial cultures were grown overnight at 18°C (~16 h), after which cells were harvested and resuspended in lysis buffer (50 mM Tris-

HCl [pH 7.5], 500 mM NaCl, 5 mM imidazole, 1 mM DTT) supplemented with 1 mg/ml lysozyme and protease inhibitor cocktail. Cells were lysed by sonication, and the supernatant was then clarified by centrifugation at 18,000 rpm for 30 min. The supernatant was incubated with preequilibrated Ni-NTA agarose (Qiagen) for 1 h. The resin was then washed twice with wash buffer (50 mM Tris-HCl [pH 7.5], 500 mM NaCl, 25 mM imidazole, 1 mM DTT). The proteins were eluted in elution buffer containing 300 mM imidazole. For Acr proteins, the 6×His tag was removed by incubation with His-tagged tobacco etch virus (TEV) protease overnight at 4°C followed by a second round of Ni-NTA purification to isolate successfully cleaved, untagged anti-CRISPRs (by collecting the unbound fraction).

In vitro DNA cleavage

For all *in vitro* DNA cleavage experiments, sgRNA targeting a PCR fragment was generated by *in vitro* T7 transcription, using the sgRNA transcription kit from NEB. Cas9 (300-500 nM) was incubated with purified, recombinant anti-CRISPR protein in cleavage buffer [20 mM HEPES-KOH (pH 7.5),150 mM KCl, 10% glycerol, 1 mM DTT, and 10 mM MgCl2] for 10 min. Next, sgRNA (1:1, 300-500 nM) was added and the mixturewas incubated at room temperature for another 15 min. DNA containing the target protospacer was generated by restriction digestion of a plasmid or generated by PCR amplification. The target DNA was added to the Cas9/sgRNA complex at 5-10 nM final concentration. The reactions were treated with 1U proteinase

K from NEB, incubated at 37C for 30 min and visualized after electrophoresis in a 1% agarose/1xTAE gel.

For the Acr titration experiment, (150 nM) was incubated with purified, recombinant anti-CRISPR protein (0 to 5 μM) in cleavage buffer (20 mM HEPES-KOH [pH 7.5], 150 mM KCl, 1 mM DTT) for 10 min. Next, sgRNA (1:1, 150 nM) was added and the mixture was incubated for another 15 min. Plasmid containing the target protospacer NTS4B was linearized by ScaI digestion. Linearized plasmid was added to the Cas9/sgRNA complex at 3 nM final concentration. The reaction mixtures were incubated at 37°C for 60 min, treated with proteinase K at 50°C for 10 min, and visualized after electrophoresis in a 1% agarose/1× TAE gel.

Cas9-Acr copurification

Cas9 proteins were expressed from plasmid pMCSG7 with an N-terminal 6×His affinity tag in *E. coli* Rosetta cells. Untagged Acrs were coexpressed in the same cells from plasmid pCDF1b. Cells were grown in LB to an OD_{600} of 0.8, and protein production was induced with 2 mM IPTG overnight at 16°C. Cells were collected by centrifugation, resuspended in binding buffer (20 mM Tris, pH 7.5, 250 mM NaCl, 5 mM imidazole), and lysed by sonication, and cellular debris was removed by centrifugation. The cleared lysates were applied to Ni-NTA columns, washed with binding buffer supplemented with 30 mM imidazole, and eluted with 300 mM imidazole. Protein complexes were analyzed by SDS-PAGE followed by Coomassie staining.

Electrophoretic mobility shift assay (EMSA)

NmeCas9 (1 μM) was incubated with 1 μM sgRNA in 1× binding buffer (20 mM Tris-HCl [pH 7.5], 150 mM KCl, 2 mM EDTA, 1 mM DTT, 5% glycerol, 50 μg/ml heparin, 0.01% Tween 20, 100 μg/ml BSA) for 20 min at room temperature to form the RNP complex. Acrs were added to a final concentration of 10 μM and incubated for an additional 20 min. Finally, the FAM-tagged NTS4B protospacer oligonucleotide was added to the mixture and incubated at 37°C for 1 h. The mixture was loaded onto a native 6% acrylamide gel, and the FAM-tagged DNA was visualized using a Typhoon imager.

CHAPTER 3: Characterization of compact Cas9 orthologs

3.1 Introduction:

As discussed in chapter 1, *in vivo* applications of some Cas9s are hindered by large size (limiting delivery by AAV vectors), off-target editing, or complex protospacer-adjacent motifs (PAMs) that restrict the density of recognition sequences in the target of interest. Table 1.1 shows all Cas9 orthologs validated for *in vivo* genome editing using AAV. In the case of SpyCas9, packaging and administration of two separate AAV vectors are required, resulting in an increased dose and reduced therapeutic efficiency. The compact orthologs discussed in chapter 1 are all amenable to all-in-one AAV delivery, but they possess longer PAMs: N_4GATT for NmeCas9, N_2GRRT for SauCas9 and N4RYAC for CjeCas9 where R and Y stands for purines and pyrimidines, respectively. Longer PAMs reduce the number of targetable sites at a given locus. This is especially important in certain cases such as editing of small targets (e.g., miRNAs), correction of mutations by base editing (Rees and Liu, 2018) or during precise editing via HDR [which is most efficient when the rewritten bases are close to the cleavage site (Gallagher and Haber, 2018)]. Because of these PAM restrictions, many editing sites cannot be targeted using all-in-one AAV vectors for *in vivo* delivery.

While there are thousands of Cas9 orthologs found in nature, only a dozen orthologs have been characterized, and a handful validated for genome editing. This is partially due to the difficulty in characterizing new orthologs and is compounded by the fact that not all orthologs are functional in human cells. Previously, Cas9 orthologs were selected randomly from known pathogens or laboratory bacteria. In the search for

a more rational approach, we utilized our knowledge of Acrs as indicators of type II-C Cas9 orthologs that are highly active. Specifically, we hypothesized that the existence of an Acr protein against a Cas9 ortholog signifies a viable and active Cas9. Here, we characterize four compact Cas9 orthologs and test their ability to provide CRISPR immunity in *E. coli* and enable genome editing in mammalian cells.

3.2 Results:

3.2.1 Selection of compact Cas9 orthologs to be characterized

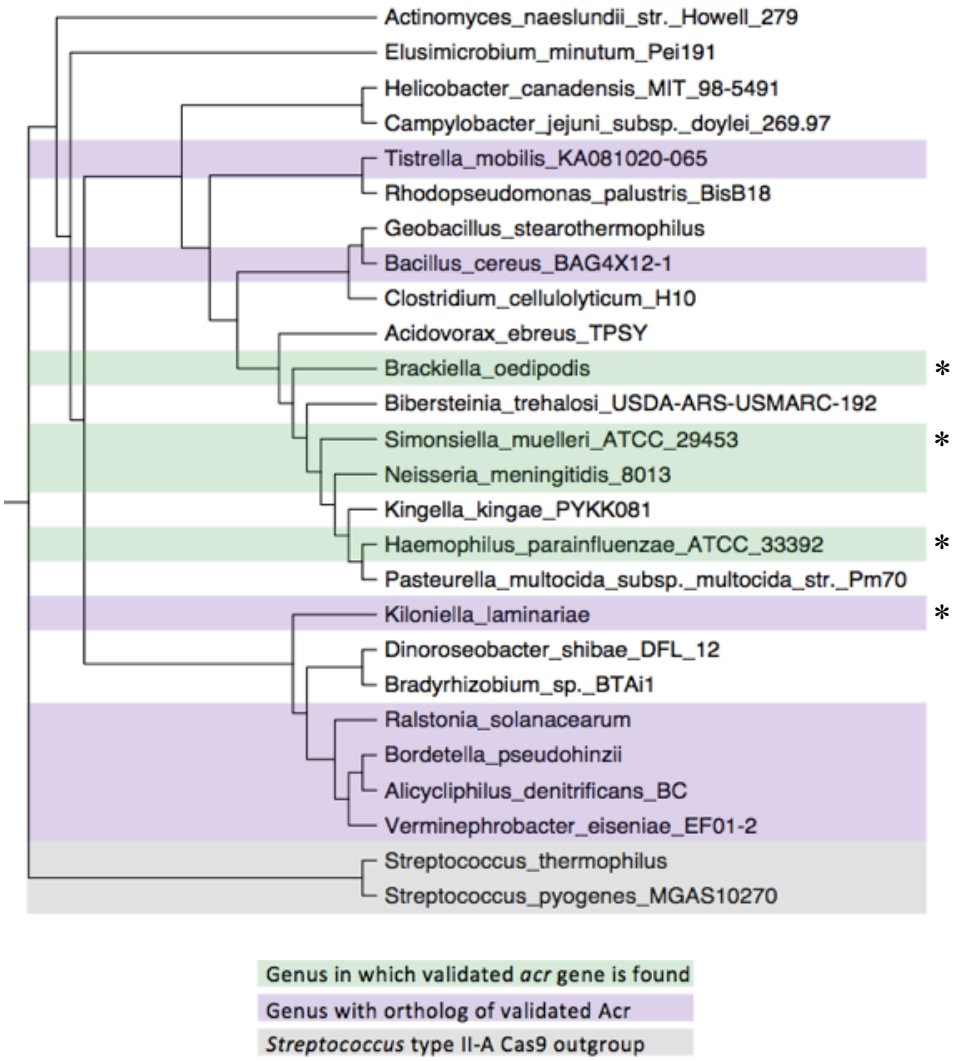
We selected four orthologs with the following properties: A) The genus to which the Cas9 belongs possesses either a validated *acr* gene or has an ortholog of a validated *acr* gene (shown in green and purple, respectively); B) The Cas9 is from bacteria that could be handled in a BSL2 lab; and C) Cas9s that are compact (>1150 aa) which would be amenable to single-vector AAV delivery. This resulted in selection of four Cas9 orthologs (indicated by * in figure 3.1A) from diverse bacteria: *Haemophilus parainfluenzae* (HpaCas9), *Simonsiella muelleri* (SmuCas9), *Brackiella oedipodis* (BoeCas9) and *Kiloniella laminariae* (KlaCas9). The size of these Cas9s and their % identity to NmeCas9 is shown in Figure 3.1B.

S. mueller and *H. parainfluenzae* are closely-related to *Neisseria* species. All three genera are found as human commensals in the nasopharynx and likely evolved from a recent common ancestor (Norskov-Lauritsen, 2014). For example, *S. muelleri* and *Neisseria dendrificans* 16S rDNAs show high similarity levels of ~96% (Hedlund and Staley, 2002). The other two orthologs come from bacteria that are much more divergent. *B. oedipodis* is a rod-shaped bacterium that infects the hearts of the endangered Cotton-Topped Tamarin (*Sanguius oedipus*) and causes endocarditis (Willems *et al.*, 2002). *K. laminariae* is an alphaproeobacterium from a marine sugar kelp (*Laminaria saccharina*). *K. laminariae* grows optimally at room temperature and lower pH (pH 5.5) and relatively high salt. It has a unique phylogenetic position and is highly diverged from other Alphaproteobacteria (Wiese *et al.*, 2009). These

characteristics suggest that the Cas9 orthologs from these bacteria may have unique properties.

In addition to Cas9, there are three other components that must be known before characterizing the activity of these unknown Cas9s. These include the crRNA, tracrRNA and the PAM. Absent any RNA-seq data available from these bacteria, we relied on bioinformatic predictions to identify the crRNA and the tracrRNA. To do so, the CRISPR loci were predicted using CRISPRfinder (Grissa, Vergnaud and Pourcel, 2007). The crRNAs were identified using the homology of the repeats to the *N. meningitdis* locus (see Appendix 1 for sequences) (Zhang *et al.*, 2013). For the tracrRNA prediction, we relied on complementarity to the constant region of the crRNA, conservation with known tracrRNAs, and the presence of a putative promoter and terminator (Figure 3.2A, HpaCas9 shown as an example, see Appendix 1 for all sgRNAs) (Lesnik *et al.*, 2001). We were able to predict a sgRNA scaffold for every Cas9 except for SmuCas9. In fact, the CRISPR-Cas locus of *S. muelleri* ATCC 29453 appears to be degenerate (Figure 3.2B). There is no apparent *cas1*, and the *cas2* lacks a canonical ATG start codon. However, the Cas9 ORF (1,065 aa) is intact and has all the predicted functional domains found in other Cas9 orthologs, which suggested that SmuCas9 itself might be active. When we attempted to define an appropriate guide RNA scaffold for SmuCas9, we could not predict its tracrRNA (based in part on crRNA complementarity) from nearby genomic sequences. Instead, we found an IS5 integrase upstream of *cas9*, where a tracrRNA locus is often observed (Figure 3.2B). Although we sequenced ~2 kb flanking the CRISPR locus to fill gaps in the genome assembly,

we could not detect a tracrRNA sequence. As an alternative, we took advantage of the nonorthogonality of sgRNAs to closely related Cas9 orthologs [guides from closely related Cas9s can be used interchangeably with some compromise in Cas9 activity (Fonfara *et al.*, 2014)] and used the NmeCas9 sgRNA to test the cleavage activity of SmuCas9.



Cas9	Host	Length (aa)	Identity to NmeCas9 (%)
Neisseria meningitidis (Nme)	Humans	1082	100
Haemophilus parainfluenzae (Hpa)	Humans	1052	66
Simonsiella muelleri (Smu)	Humans	1065	62
Brackiella oedipodis (Boe)	Cotton-Topped Tamarin	1148	46
Kiloniella laminariae (Kla)	sugar kelp	1091	28

Figure 3.1 Selection of compact Cas9 orthologs to be characterized
Phylogenetic tree of several compact Cas9 orthologs, * indicated those selected for further characterization.

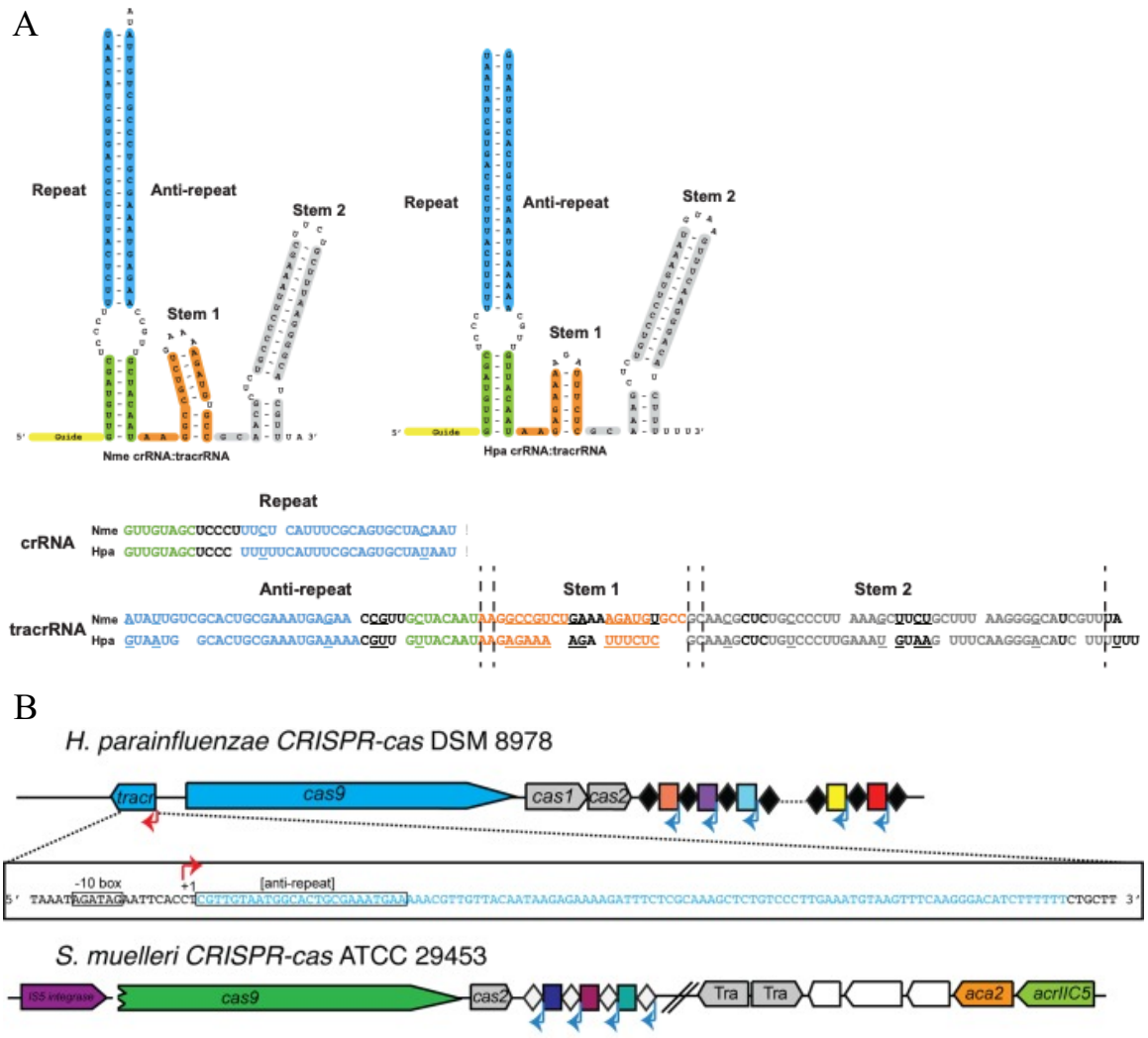


Figure 3.2 Characterization of new type II-C Cas9 orthologs.
(A) Predicted crRNA:tracrRNA structures for NmeCas9 and HpaCas9. Nucleotides that are different between the two orthologs are underlined.
(B) Genomic architectures of CRISPR-cas loci of *H. parainfluenzae* DSM 8978 and *S. muelleri* ATCC 29453. The sequence of the HpaCas9 tracrRNA is shown in the inset. Individual genomic elements are not drawn to scale.

3.2.2 The selected Cas9 orthologs recognize unique PAM sequences

PAMs are often difficult to predict and in most cases, PAMs must be identified empirically. To identify the PAMs of our selected Cas9 orthologs, we used *in vitro* PAM identification (Kim et al., 2017). The method consists of using recombinant Cas9, along with an *in vitro* transcribed sgRNA that targets a pool of DNA fragments containing a protospacer followed by a 10-bp randomized sequence cleaved *in vitro* (Figure 3.3A). The reaction products are then gel purified followed by library preparation and deep sequencing. We followed a modified deep sequencing protocol designed for strand-specific libraries (Zhang *et al.*, 2012). We used SpyCas9 as a positive control and retrieved the canonical NGG PAM, as expected (Figure 3.3B). The analysis of cleaved fragments revealed the PAMs of each Cas9 ortholog: N_4GATTT for HpaCas9, N_4C for SmuCas9 (using NmeCas9 sgRNA), N_4CC for BoeCas9 and NGGGGA for KlaCas9. To the best of our knowledge, the PAM of SmuCas9 is the shortest PAM ever recorded for a naturally occurring Cas9. Conversely, HpaCas9 and KlaCas9 have the longest PAMs observed to date.

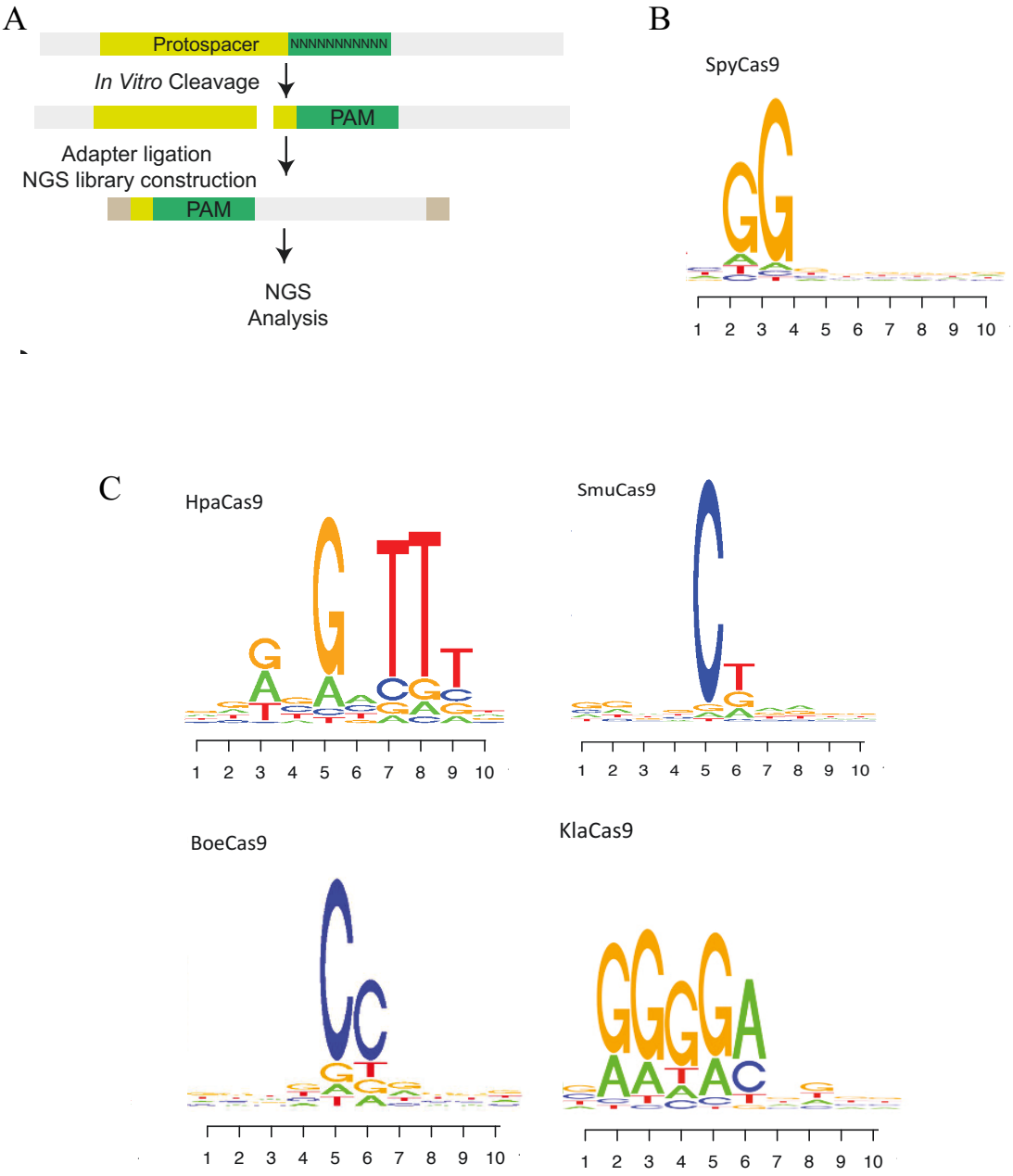


Figure 3.3 The new Cas9 orthologs recognize diverse PAMs
(A) Experimental workflow of the *in vitro* PAM discovery assay with a 10-bp randomized PAM region. Following *in vitro* digestion, adapters were ligated to cleaved products for library construction and sequencing.
(B and C) Sequence logos resulting from *in vitro* PAM discovery (B) reveal the expected NGG PAM for SpyCas9 control and (C) the PAMs of the new Cas9s.

3.2.3 Heterologous expression of the characterized Cas9s provides immunity

After our initial biochemical characterization, we wanted to establish an *in vivo* activity assay to test the ability of each ortholog to interfere with phage infections. Previous work has shown that heterologous expression of a CRISPR system in another host results in efficient CRISPR interference (Sapranauskas *et al.*, 2011), given the mobile nature of CRISPR systems acquired through horizontal gene transfer or plasmids (Makarova *et al.*, 2011). In fact, plasmid-based expression of NmeCas9 *in E. coli* with a sgRNA targeting another plasmid results in efficient interference (not shown). NmeCas9 also provides immunity against phages in *E. coli*.

To test our new Cas9 orthologs *in vivo*, we cloned each Cas9 into bacterial expression vectors along with sgRNAs targeting phage *Mu* (Morgan *et al.*, 2002). We used NmeCas9 as a positive control. A non-targeting sgRNA leads to efficient plaque formation by the phage, as evident in the plaque assay shown in Figure 3.4. In contrast, plasmid-mediated expression of HpaCas9, BoeCas9 and KlaCas9 (at 30 °C, see below) and an sgRNA designed to target *E. coli* phage *Mu* leads to significant reduction in plaque formation, suggesting highly active CRISPR interference. We did not detect any significant interference with SmuCas9, perhaps due to the use of the noncognate sgRNA from NmeCas9 (not shown).

It is important to note that our initial attempts at performing this assay using KlaCas9 at 37C were unsuccessful (not shown). We reasoned that the aquatic origin of KlaCas9 may mean a lower optimal temperature, although it functioned at 37 °C *in*

vitro. These results suggest that KlaCas9 may not be ideal for mammalian genome editing.

Cas9	Non-targeting	Targeting
NmeCas9		
HpaCas9		
BoeCas9		
KlaCas9		

Figure 3.4 The new CRISPR-Cas9 orthologs provide immunity against phage *Mu*
Representative phage *Mu* plaque assays for each of the Cas9 systems tested. 10-fold serial dilution of phage lysate were plated on *E. coli* lawns expressing each Cas9. The non-targeting sgRNA allows robust phage plaquing, while the phage *Mu* targeting sgRNA mediates strong inhibition of phage plaquing.

3.2.4 The search for an optimal SmuCas9 sgRNA

The apparent single-nucleotide PAM of SmuCas9 makes it a promising candidate with high target site density for human genome editing applications. However, our initial results revealed that although SmuCas9 is functional with the sgRNA from NmeCas9 *in vitro*, it fails to efficiently cleave DNA in *E. coli.* We hypothesized that sgRNA incompatibility (i.e. low affinity between SmuCas9 and Nme sgRNA in cells) may be the reason for this inefficiency. To address this, we sought to identify an sgRNA from an ortholog that is more similar to SmuCas9 with the potential to improve complex formation for genome editing. A BLASTN search revealed that no other *Simonsiella* species has been sequenced, and the most closely related Cas9s are only ~70% identical (compared to ~60% with NmeCas9). Nevertheless, we selected the locus from *Bergeriella dendrificans* (BdeCas9, 70% identity to SmuCas9) due to an easily identifiable tracrRNA. We first performed *an in vitro* cleavage using SmuCas9 with *in vitro* transcribed Bde and Nme sgRNAs to compare the kinetics of cleavage. As expected, SmuCas9 more efficiently cleaves a target *in vitro* with the Bde sgRNA (not shown).

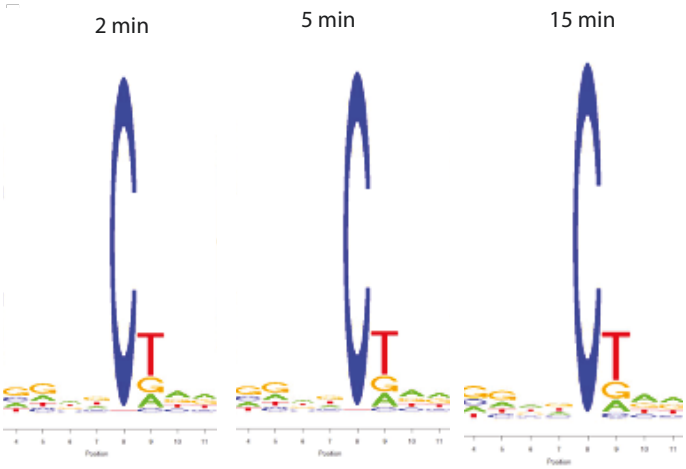
Next, we performed the *in vitro* PAM discovery assay using SmuCas9-Bde sgRNA to uncover any additional nucleotides that might be required. To capture the most efficient PAMs for cleavage, we performed the assay at the lowest possible concentration possible for a signal (30nM Cas9-sgRNA) and we stopped the reaction 2,

5 and 15 minutes after the DNA was added. This revealed a slight bias for a T at position 6, which may be important for efficient targeting in cells.

A

WP_002641950.1	[Simonsiella muelleri]	100%
WP_085365262.1	[Neisseria dentiae]	71%
WP_066076886.1	[Bergeriella denitrificans]	70%
WP_085358855.1	[Neisseria dumasiana]	70%
WP_054599015.1	[Neisseria sp. 74A18]	70%
WP_085355450.1	[Neisseria dumasiana]	70%
WP_007341035.1	[Neisseria bacilliformis]	65%
WP_049256670.1	[Neisseria bacilliformis]	65%
WP_054618016.1	[Neisseria sp. 83E34]	66%
WP_067593048.1	[Eikenella sp. NML02-A-017]	63%
WP_049229472.1	[Neisseria lactamica]	63%
WP_101126884.1	[Neisseria meningitidis]	63%
WP_101116244.1	[Neisseria meningitidis]	63%

B



C

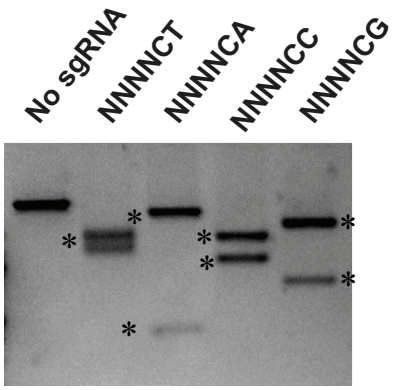


Figure 3.5 SmuCas9 cleaves N_4C PAMs *in vitro* with Bde sgRNA
(A) BLASTN results showing the most closely related Cas9s to SmuCas9.
(B) *In vitro* PAM identification (same as Figure 3.3A) using SmuCas9-BdeSgRNA at 3 timepoints to identify possible preferences other than the C at the 5th position.
(C) *In vitro* cleavage of protospacers bearing N_4CA, N_4CG, N_4CC and N_4CT PAMs. Cleaved bands are marked with *.

To test the ability of SmuCas9 to cleave sites with N4CN PAMs *in vitro*, we performed *in vitro* cleavage on substrate with N_4CA, N_4CG, N_4CC and N_4CT PAMs. We used standard concentrations of Cas9 (30nM) vs target (10nM) that we used previously for our other *in vitro* studies, as high concentrations may mask potential preference (Figure 3.5C). The results suggest that SmuCas9 is capable of efficiently cleaving N_4CN PAMs *in vitro* using the sgRNA from *Bergeriella dendrificans* and can be used to test SmuCas9 in mammalian genome editing.

3.2.5 Genome editing with the characterized Cas9 orthologs

To test the ability of the characterized Cas9s in editing the human genome, we cloned each Cas9 from bacterial gDNA into a mammalian expression plasmid previously used for NmeCas9 (containing NLSs and appropriate promotors). However, transfection of these into HEK293T cells showed that all Cas9 orthologs were expressed poorly (not shown), possibly due to the fact that they are not codon optimized for mammalian cells.

Several studies have used a ribonucleoprotein (RNP) consisting of recombinant Cas9 and the sgRNA into cells for robust genome editing, circumventing the need for expression from a plasmid. We cloned these Cas9s along with NLSs into a bacterial expression vector, recombinantly expressed and purified them. We electroporated HEK293T cells with a preformed RNP complex of each Cas9 along with *in vitro* transcribed sgRNAs. We designed at least two sgRNAs for each ortholog to account for the possibility of a nonfunctional sgRNA, targeting several loci throughout the genome with the appropriate PAMs (table 3.1). Considering that SmuCas9's PAM is potentially only 1 nucleotide long, we designed 10 guides to test several combinations. 72 hours after electroporation, we isolated gDNA from cells, amplified the target locus and performed TIDE analysis. Our results showed that HpaCas9 efficiently induced indels at two of the four sites tested. While we had initially postulated that its PAM was identical to that of NmeCas9 (N_4GATT), we only observed editing at N_4GATT<u>T</u>.

However, more careful analysis is required and more sites must be tested to confirm this observation. SmuCas9 was functional at 2 of the 10 sites tested, albeit at a very low efficiency of 1-1.5%. We attribute this low efficiency to the lack of a cognate sgRNA. However, we did not detect indels by KlaCas9 and BoeCas9, suggesting they are not active in our experimental setup. KlaCas9's low efficiency could be attributed to poor activity at temperatures used to grow mammalian cells.

Name	Locus	Target protospacer	PAM	Editing (%)
SmuTS1	*AAVS1*	TCCTCCTTCCTAGTCTCCTGATAT	TCGTCTAA	0
SmuTs2	*AAVS1*	CAAAATCAGAATAAGTTGGTCCTG	AGTTCTAA	0
SmuTS3	*AAVS1*	GGGAGACATCCGTCGGAGAAGG	CCATCCTA	0
SmuTS4	*mTLR*	ATCACCTGCCTCGTGGAATACGGTAAACCTA	TAAACCTA	0
SmuTS5	*mTLR*	TCACCTGCCTCGTGGAATACGGTAACCTAC	AAACCTAC	1.5
SmuTS6	*mTLR*	GAGACAAATCACCTGCCTCGTGGAATACGGT	AATACGGT	0
SmuTS7	*mTLR*	CCGGCTTTGGCGAGACAAATCACCTGCCTCG	CTGCCTCG	1
SmuTS8	*mTLR*	GTTCAGCGTGTCCGGCTTTGGCGAGACAAA	GAGACAAA	0
SmuTS9	*mTLR*	TTCCACGAGGCAGGTGATTTGTCTCGCCAAA	TCGCCAAA	0
SmuTS10	*mTLR*	ATTCCACGAGGCAGGTGATTTGTCTCGCCAA	CTCGCCAA	0
HpaTS1	*LINC01588*	GGACAGGAGTCGCCAGAGGCCGGT	GGTGGATTT	35
HpaTsS2	*CYBB*	GCTGGATTACTGTGTGGTAGAGGG	AGGTGATTA	0
HpaTS3	*AAVS1*	TTTGCCTGGACACCCCGTTCTCC	TGTGGATTC	0
HpaTS4	*mTLR*	TAGGTTTACCGTATTCCACGAGGC	AGGTGATTT	24
BoeTS1	*AAVS1*	GGGAGACATCCGTCGGAGAAGG	CCATCCTA	0
BoeTS2	*mTLR*	ATTCCACGAGGCAGGTGATTTGTCTCGCCAA	CTCGCCAA	0
BoeTS3	*mTLR*	ATCACCTGCCTCGTGGAATACGGTAAaccta	TAAACCTA	0
KlaTS1	*AAVS1*	TATCCAAGGTTAAGCAAAAGAG	AGGGGA	0
KlaTS2	*AAVS1*	CTTGGCAGGGGGTGGGAGGGAAGG	GGGGGA	0

Table 3.1 List of sites targeted by each Cas9 ortholog

Although HpaCas9's PAM is relatively long, this could offer higher accuracy at sites that are targetable. The efficient editing by HpaCas9 allowed us to test the ability of all previously discovered Acrs in preventing genome editing by HpaCas9. Based on our *in vitro* results in Figure 2.5, we expected all type II-C Acrs to inhibit HpaCas9 genome editing. To address this question, we delivered a preformed RNP complex of HpaCas9, sgRNA, and each Acr to HEK293T cells by electroporation. Then, we confirmed genome editing inhibition using TIDE (Figure 3.5). Acrs displayed variations in activities even with RNP delivery, suggesting differences in protein stability, off-rate, or other intrinsic properties. Of note, however, AcrIIC1, AcrIIC2 and AcrIIC4 exhibited strong inhibitory potency against HpaCas9 in HEK293T cells and can be used as off-switches for HpaCas9 genome editing.

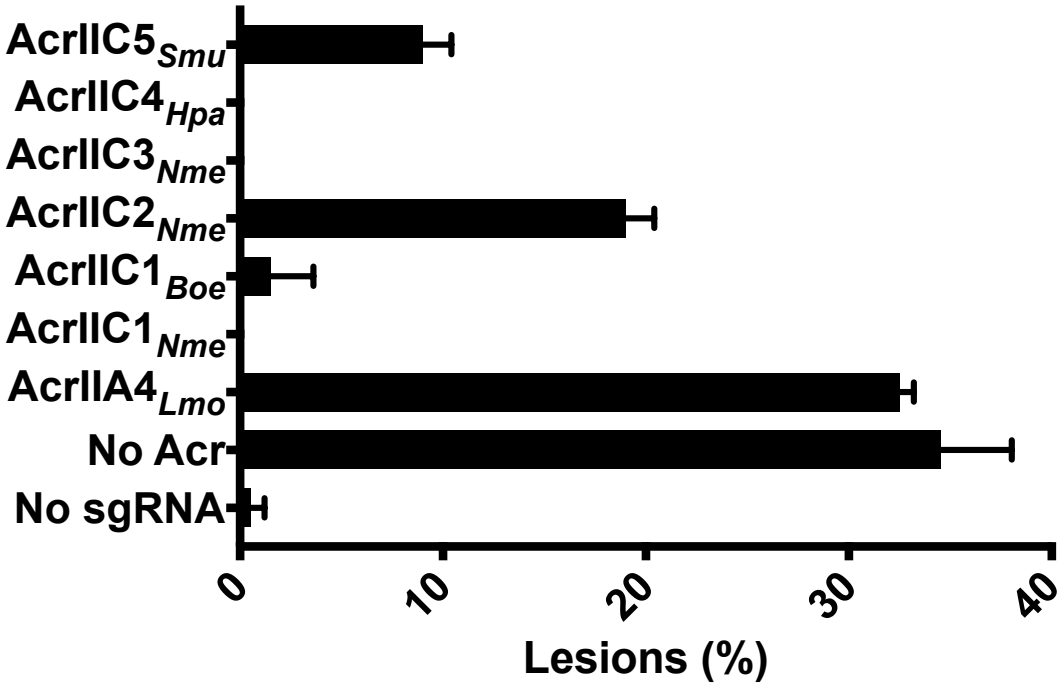


Figure 3.5: Genome editing with HpaCas9 in the presence of Acrs

Indels (%) of editing efficiencies measured by TIDE analysis upon RNP delivery of HpaCas9-sgRNA and each Acr into HEK293T cells. AcrIIA4 was used as a negative control, which is specific to type II-A CRISPR systems (Rauch *et al.*, 2017) and would not be expected to inhibit type II-C systems.

3.3 Discussion

In this chapter, we characterized compact Cas9 orthologs and tested them for genome editing applications. Instead of a random approach, we used Acr proteins as indicators for Cas9 orthologs that have higher activities. In fact, while all Cas9s worked *in vitro* and demonstrated a wide diversity of PAM recognition, only one (HpaCas9) functions in human cells. This is in line with previous efforts to characterize new Cas9 systems for genome editing (Ran *et al.*, 2015), suggesting that not every active Cas9 ortholog is functional in human cells. It is currently not possible to predict whether a novel Cas9 will be functional for editing in human cells or not, and although our Acr-directed approach yielded Cas9s that functioned *in vitro*, most were unsuccessful for editing in HEK293T cells.

The short PAM of SmuCas9 and the non-zero, albeit very low editing makes it of interest for future efforts. We hypothesize that the low editing efficiency can be attributed to the Bde sgRNA's loading/affinity for SmuCas9. Sequencing of more *Simonsiella* genomes may make possible to find a closely-related ortholog that harbors the right sgRNA for SmuCas9.

HpaCas9's N_4GNTTT PAM is, to our knowledge, the longest PAM of any Cas9 reported for genome editing (5 nucleotides long). Despite the fact that the long PAM limits the number of targetable sites in the human genome (even more so than for other compact Cas9s), it may target such sites with very high accuracy, similar to the extreme accuracy that has been documented for NmeCas9.

3.4 Materials and methods:

Plasmids and construct cloning

For genome editing experiments, plasmids expressing NmeCas9, SpyCas9 and their respective sgRNAs targeting the DTS3 site, as well as plasmids expressing AcrE2 (Addgene #85677), AcrIIC1Boe (Addgene #85678) and AcrIIC1Nme (Addgene #85679), were previously described (Pawluk et al.,2016a, 2016b). AcrIIA4 expressing plasmid (Addgene #86842) was described previously (Rauch et al., 2017). The CjeCas9-expressing plasmid (PX404) was acquired from Addgene (#68338). For CjeCas9 sgRNA expression, the published sgRNA sequence (Kim et al., 2017) was synthesized as a gBlock (IDT), and was used to replace the NmeCas9 sgRNA cassette in pLKO.1-puro plasmid[(Pawluk et al., 2016b); Addgene #85715] by Gibson Assembly. The resulting plasmid (pEJS676) contains the CjeCas9 sgRNA cassette with BfuAI sites that can be used to insert any spacer of interest. Next, two previously validated guide sequences [targeting the AAVS1 locus (TS2 and TS6; (Kim et al., 2017)] were inserted into the CjeCas9 sgRNA expression construct, yielding plasmids pEJS677 and pEJS678, respectively.

HEK293T electroporation and indel analysis

HpaCas9 was delivered into HEK293T cells as RNP complex using Neon electroporation system. 2uL of Purified HpaCas9 protein [10uM] was complexed with 1uL of 3.3Ug/uL gRNA. When Acr is tested, 2uL of purified Acr proteins [50uM] was added. Electroporation parameters (voltage, width, number of pulses) were 1150 V, 20 ms,

2 pulses. Electroporated cells were plated in 24-well plate and incubated for 48 hours before cells were harvested and gDNA was extracted. Indel analysis was performed by TIDE.

Purification of Cas9

6xHis-Cas9:sgRNA was expressed in *E. coli* Rosetta (DE3). Cells were grown in Terrific Broth (TB) medium at 37 °C to an optical density (OD600 nm). Protein expression was induced by the addition of 1 mM IPTG for 16 h at 16 °C. Cells were lysed by sonication in 50 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.5 mM DTT and 5% glycerol supplemented with 0.5 mM PMSF, lysozyme and protease inhibitor cocktail (Sigma). Clarified lysates were bound in batch to Ni-NTA agarose (Qiagen), and bound protein was eluted with 300 mM imidazole. Purified Cas9:sgRNA was dialyzed into 20 mM HEPES pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM DTT and 1 mM PMSF) for protein interaction experiments.

In vitro DNA cleavage

NmeCas9 sgRNA derived from spacer 25 (Zhang et al., 2015) was generated by *in vitro* T7 transcription (Epicentre). NmeCas9 (500 nM) was incubated with purified, recombinant anti-CRISPR protein in cleavage buffer [20 mM HEPES-KOH (pH 7.5), 150 mM KCl, 10% glycerol, 1 mM DTT, and 10 mM MgCl2] for 10 minutes. Next, sgRNA (1:1, 500 nM) was added and the mixture was incubated for another 15 minutes. Plasmid containing the target protospacer 25 (pEJS560) was linearized by

ScaI digestion. Linearized plasmid was added to the Cas9/sgRNA complex at ~5 nM final concentration. The reactions were incubated at 37 °C for 30 minutes and visualized after electrophoresis in a 1% agarose/1xTAE gel.

In vitro PAM determination

A library of a protospacer with randomized PAM sequences was generated using overlapping PCRs, with the forward primer containing the 10-nt randomized sequence flanking the protospacer. The library was subjected to *in vitro* cleavage by purified recombinant HpaCas9 or SmuCas9 proteins as well as *in vitro*-transcribed sgRNAs. Briefly, 300 nM Cas9:sgRNA complex was used to cleave 300 nM target fragment in 1× reaction buffer (NEBuffer 3.1) at 37°C for 60 min. The reaction mixture was then treated with 1 U proteinase K (NEB) at 50°C for 10 min and run on a 4% agarose gel with 1× TAE. The segment of a gel where the cleavage products were expected to be was purified and subjected to library preparation as described previously (Zhang *et al.*, 2012). The library was sequenced using the Illumina NextSeq500 sequencing platform and analyzed with custom scripts.

CHAPTER 4: Nme2Cas9 as a compact, accurate Cas9 with a dinucleotide PAM

4.1 Introduction:

As mentioned previously, *in vivo* applications of some Cas9s are hindered by large size, off-target editing, or complex PAMs that restrict the density of recognition sequences in target DNA. Our efforts in chapter 3 yielded several compact Cas9 orthologs with novel PAMs, but most were unable to execute efficient genome editing in mammalian cells, or had long PAMs, or both.

We continued our search for an ideal Cas9 ortholog for single-AAV delivery using an evolutionary approach. PAM mutations often enable phages to escape from type II immunity (Paez-Espino *et al.*, 2015), placing these systems under selective pressure not only to acquire new CRISPR spacers, but also to evolve new PAM specificities. In addition, as shown in chapter 2, some phages and MGEs express Acr proteins that inhibit Cas9, sometimes by PID interactions that could confer selective pressure for PID variation (Pawluk, Davidson and Maxwell, 2018; Shin *et al.*, 2017; Dong *et al.*, 2017). Cas9 PIDs can evolve such that closely related orthologs recognize distinct PAMs, as illustrated recently in two species of *Geobacillus* (Harrington et al., 2017).

We exploited natural variation in the PAM-interacting domains (PIDs) of closely related Cas9s to identify a compact ortholog from *Neisseria meningitidis*, Nme2Cas9, that recognizes a simple dinucleotide PAM (N_4CC) that provides for high target site density. All-in-one AAV delivery of Nme2Cas9 with a guide RNA targeting

Pcsk9 in adult mouse liver produces efficient genome editing and reduced serum cholesterol with exceptionally high specificity. We further expand our single-AAV system to pre-implanted zygotes for streamlined generation of genome-edited mice. Finally, we show the crystal structure of Nme2Cas9 and the mechanism of PAM recognition.

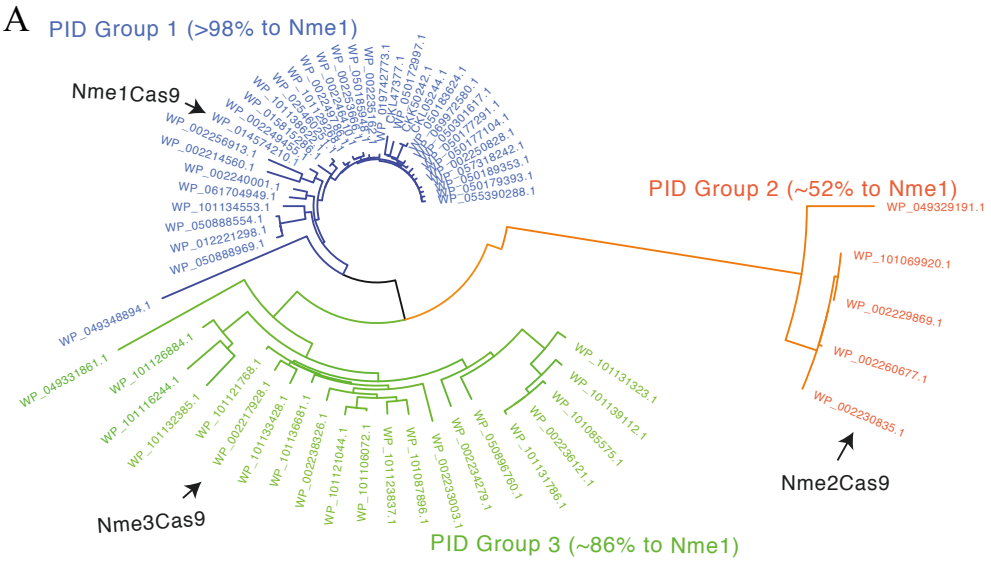
4.2 RESULTS

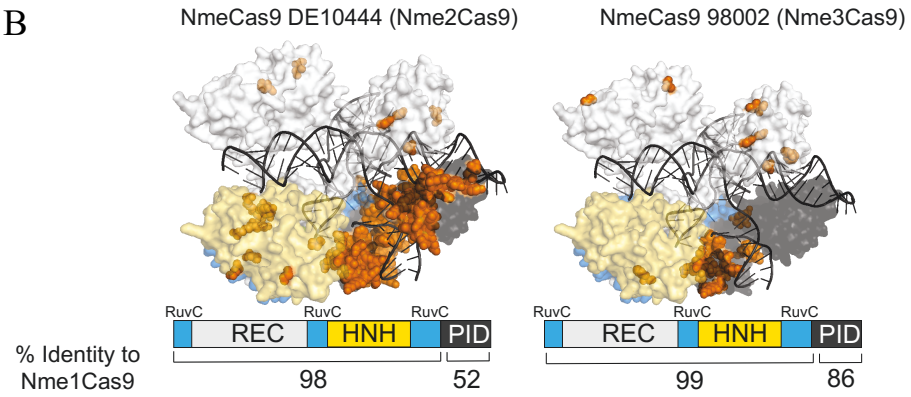
4.2.1 Closely related orthologs of NmeCas9 recognize different PAMs

We and others have shown that NmeCas9 is an efficient and accurate Cas9 for genome editing applications. Given that numerous *N. meninigitidis* strains are sequenced, we sought closely related NmeCas9 orthologs with divergent PIDs. We focused on Cas9s with >80% overall identity to that of strain 8013 (Zhang *et al.*, 2013), which we will hereafter refer to as Nme1Cas9. Alignments revealed three clades of orthologs, each with >98% identity in the N-terminal ~820 amino acid (aa) residues, which include all regions of the protein other than the PID (Figure 4.1A). All of these Cas9s are 1,078–1,082 aa in length. The first clade (group 1) includes orthologs in which the >98% aa sequence identity with Nme1-Cas9 extends through the PID. In contrast, the other groups had PIDs that were significantly divergent from that of Nme1- Cas9, with group 2 and group 3 orthologs averaging 52% and 86% PID sequence identity with Nme1Cas9, respectively (Figure 4.1B). We selected one meningococcal strain from each group —De11444 from group 2 and 98002 from group 3— for detailed analysis, and we refer to their Cas9 orthologs as Nme2Cas9 (1,082 aa) and Nme3Cas9 (1,081 aa), respectively (see Appendix for sequences). The CRISPR-Cas loci from these two strains (Figure 4.1 C) have repeat sequences and spacer lengths identical to those of strain 8013, suggesting that their mature crRNAs also have 24-nt guide sequences and 24-nt repeat sequences (Zhang et al., 2013). Similarly, the tracrRNA sequences of De11444 and 98002 were 100% identical to the 8013 tracrRNA (Figure 4.1C). These observations imply that the same sgRNA sequence scaffold can

guide all three Cas9s and we will likely not encounter problems regarding sgRNA prediction seen in chapter 3.

To test whether these Cas9 orthologs have distinct PAMs, we replaced the PID of Nme1Cas9 with that of either Nme2Cas9 or Nme3Cas9 and used these recombinant chimeras for *in vitro* PAM identification described in chapter 3. The expected N_4GATT consensus was recovered for Nme1Cas9, validating our workflow (Figure 4.2A). The PID-swapped derivatives strongly preferred a C in the 5th position (Figure 4.2B), but other PAM nucleotides could not be confidently assigned due to the low cleavage efficiencies of the chimeric proteins under the conditions used (Figure 4.2C). To further resolve the PAMs, we repeated the *in vitro* assay but on a 7-nt randomized library possessing an invariant C at the 5th PAM position (NNNNCNNN). This strategy yielded a much higher cleavage efficiency (Figure 4.2C), and the results indicated that the Nme2Cas9 and Nme3Cas9 PIDs recognize NNNNCC(a) and NNNNCAAA PAMs, respectively (Figure 4.2D). Finally, we repeated our tests using the full-length Nme2Cas9 with the NNNNCNNN DNA pool, and again we recovered a NNNNCC(a) consensus (Figure 4.2D), though with more efficient cleavage (Figure 4.2C). These data suggest that one or more of the 15 aa changes in Nme2Cas9 (relative to Nme1Cas9) outside of the PID support efficient activity (Figure 4.2C). Because the 2- to 3-nucleotide PAM of Nme2Cas9 affords a higher density of potential target sites than the previously described compact Cas9 orthologs, we selected it for further analyses.


A
PID Group 1 (>98% to Nme1)
Nme1Cas9
PID Group 2 (~52% to Nme1)
Nme2Cas9
Nme3Cas9
PID Group 3 (~86% to Nme1)


B
NmeCas9 DE10444 (Nme2Cas9)
NmeCas9 98002 (Nme3Cas9)
RuvC
REC
HNH
PID
% Identity to Nme1Cas9
98
52
99
86

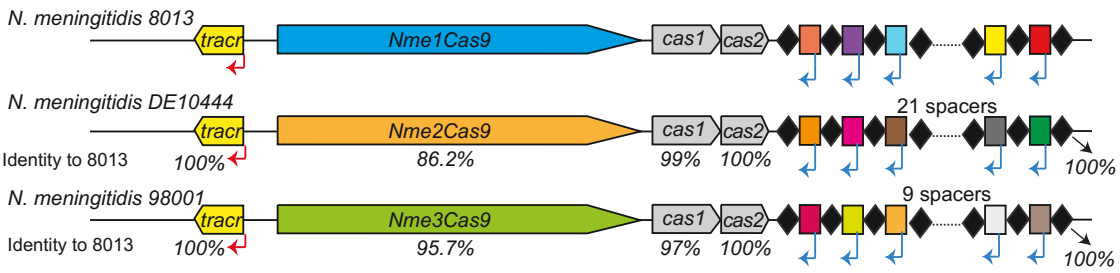

C
N. meningitidis 8013
tracr
Nme1Cas9
cas1
cas2
N. meningitidis DE10444
Nme2Cas9
21 spacers
Identity to 8013
100%
86.2%
99%
100%
N. meningitidis 98001
Nme3Cas9
9 spacers
95.7%
97%

Figure 4.1 Three closely related *Neisseria meningitidis* Cas9s have distinct PIDs
(A) Unrooted phylogenetic tree of meningococcal Cas9 orthologs. Groups 1 (blue), 2 (orange), and 3 (green) have PIDs with >98%, 52%, and 86% identity to Nme1Cas9, respectively. Three representative Cas9 orthologs (one from each group) (Nme1Cas9, Nme2Cas9, and Nme3Cas9) are indicated.
(B) Schematic showing mutated residues (orange spheres) between Nme2Cas9 (left) and Nme3Cas9 (right) mapped onto the predicted structure of Nme1Cas9, revealing the cluster of mutations in the PID (black).
(C) Schematic showing the CRISPR-Cas loci of the strains encoding the three Cas9 orthologs (Nme1Cas9, Nme2Cas9, and Nme3Cas9) from (A). Percent identities of each CRISPR-Cas component with N. meningitidis 8013 (encoding Nme1Cas9) are shown. Blue and red arrows denote pre-crRNA and tracrRNA transcription initiation sites, respectively.

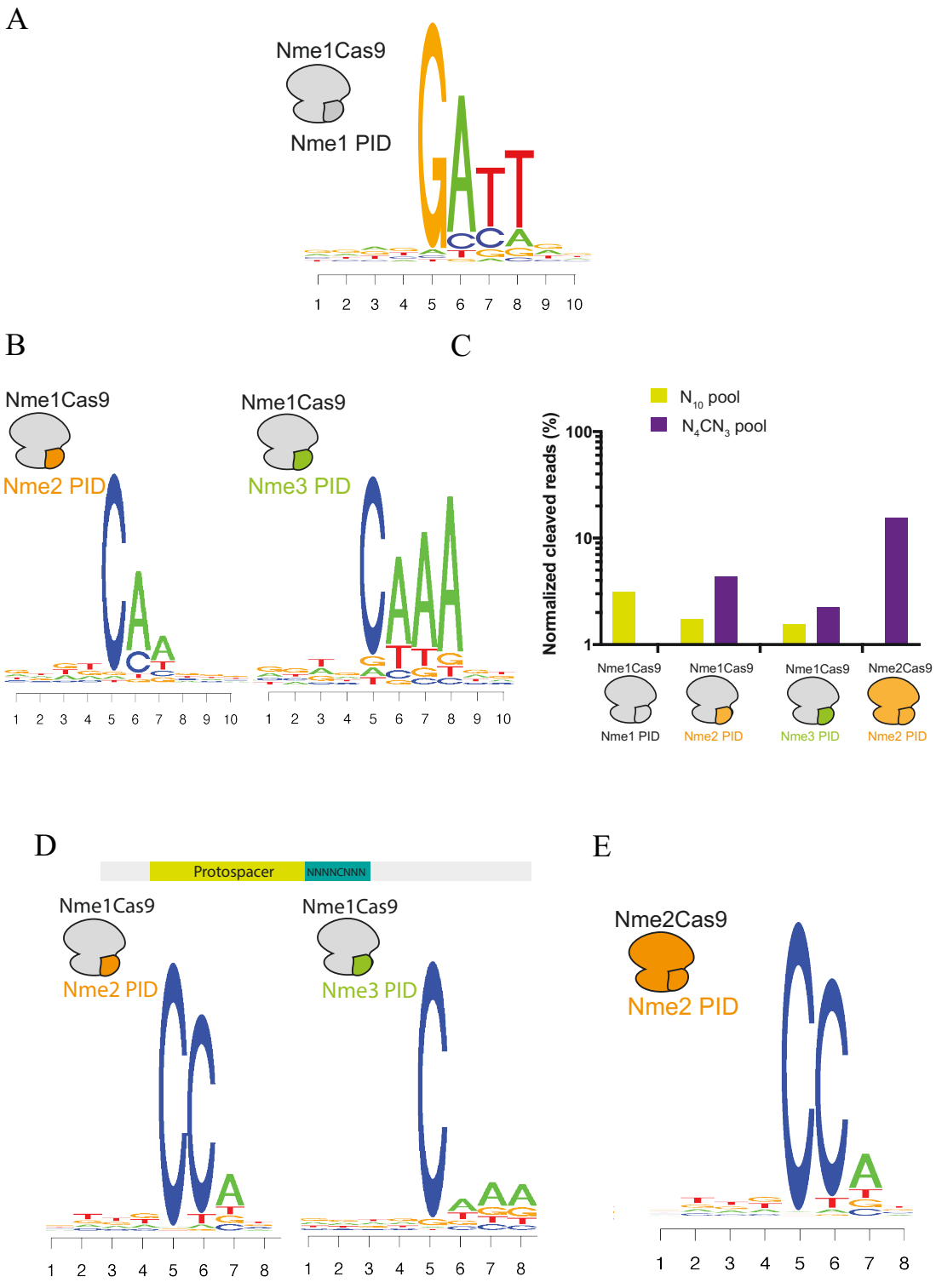

A
Nme1Cas9
Nme1 PID
1 2 3 4 5 6 7 8 9 10
B
Nme1Cas9
Nme2 PID
1 2 3 4 5 6 7 8 9 10
Nme1Cas9
Nme3 PID
1 2 3 4 5 6 7 8 9 10
C
N10 pool
N4CN3 pool
Normalized cleaved reads (%)
100
10
1
Nme1Cas9
Nme1Cas9
Nme1Cas9
Nme2Cas9
Nme1 PID
Nme2 PID
Nme3 PID
Nme2 PID
D
Protospacer
NNNNCNNN
Nme1Cas9
Nme2 PID
1 2 3 4 5 6 7 8
Nme1Cas9
Nme3 PID
1 2 3 4 5 6 7 8
E
Nme2Cas9
Nme2 PID
1 2 3 4 5 6 7 8

Figure 4.2 Three closely related Cas9s recognize different PAMs
(A) Sequence logos resulting from *in vitro* PAM discovery reveal the enrichment of a N_4GATT PAM for Nme1Cas9, consistent with its previously established specificity.
(B) Sequence logos indicate that Nme1Cas9 with its PID swapped with that of Nme2Cas9 (left) or Nme3Cas9 (right) requires a C at PAM position 5. The remaining nucleotides were not determined with high confidence due to the modest cleavage efficiency of the PID-swapped protein chimeras (see figure 4.D)
(C) Normalized read counts (% of total reads) from cleaved DNAs from the *in vitro* assays for intact Nme1Cas9 (grey), for chimeras with Nme1Cas9's PID swapped with those of Nme2Cas9 and Nme3Cas9 (mixed colors), and for full-length Nme2Cas9 (orange), are plotted. The reduced normalized read counts indicate lower cleavage efficiencies in the chimeras.
(D) Sequence logos from the *in vitro* PAM discovery assay on an NNNNCNNN PAM pool by Nme1Cas9 with its PID swapped with those of Nme2Cas9 (left) or Nme3Cas9 (right).
(E) Sequence logo from the *in vitro* PAM discovery on an NNNNCNNN PAM using full-length Nme2Cas9

4.2.2 Nme2Cas9 editing at N4CC PAMs in human cells

To test the efficacy of Nme2Cas9 in human genome editing, we cloned full-length (not PID-swapped), human-codon-optimized Nme2Cas9 into a mammalian expression plasmid, with the same appended nuclear localization signals (NLSs) and linkers validated previously for Nme1Cas9 in chapters 2 and 3 (Amrani *et al.*, 2018a). For our initial tests, we used a modified, fluorescence-based Traffic Light Reporter (Certo *et al.*, 2011). This reporter consists of a disrupted GFP followed by an out-of-frame T2A peptide and mCherry cassette. When DNA double-strand breaks (DSBs) are introduced in the broken-GFP cassette, a subset of non-homologous end joining (NHEJ) repair events leave +1-frameshifted indels, placing mCherry in frame and yielding red fluorescence that can be easily quantified by flow cytometry (Figure 4.3A). [Homology-directed repair (HDR) outcomes can also be scored simultaneously by including a DNA donor that restores the functional GFP sequence, yielding green fluorescence (Certo et al., 2011).] Because some indels do not introduce a +1 frameshift, the fluorescence readout generally provides an underestimate of the true editing efficiency. Nonetheless, the speed, simplicity, and low cost of the assay makes it useful as an initial, semi-quantitative measure of genome editing in HEK293T cells carrying a single TLR2.0 locus incorporated via lentivector.

For our initial tests, we transiently co-transfected Nme2Cas9 plasmid with one of fifteen sgRNA plasmids carrying spacers that target TLR2.0 sites with N_4CC PAMs. No HDR donor was included, so only NHEJ-based editing (mCherry) was scored. Most

sgRNAs were in G23 format (i.e. a 5'-terminal G to facilitate transcription, followed by a 23nt guide sequence], as used routinely for Nme1Cas9 (Ibraheim *et al.*, 2018b). No sgRNA and an sgRNA targeting an N4GATT PAM were used as negative controls, and SpyCas9+sgRNA and Nme1Cas9+sgRNA co-transfections (targeting NGG and N_4GATT protospacers, respectively) were included as positive controls. Editing by SpyCas9 and Nme1Cas9 was readily detectable (~28% and 10% mCherry, respectively) (Figure 4.3B). For Nme2Cas9, all 15 sgRNAs targeting sites with N4CC PAMs were functional, though to various extents ranging from 4% to 20% mCherry. These fifteen sites include examples with each of the four possible nucleotides in the 7th PAM position (after the CC dinucleotide), indicating that the slight preference for an A residue observed *in vitro* (Figure 4.1B) does not reflect a PAM requirement for editing applications in human cells. The N_4GATT PAM control yielded mCherry signal similar to no-sgRNA control (Figure 4.3B).

To determine whether both C residues in the N_4CC PAM are important for editing, we also tested a series of N_4DC (D = A, T, G) and N_4CD PAM sites in TLR2.0 reporter cells (Figure 4.3C and 4.3D). We found no detectable editing at any of these sites, providing an initial indication that both C residues of the N_4CC PAM consensus are required for efficient Nme2Cas9 activity.

To further corroborate the dinucleotide PAM, we took advantage of an eGFP knock-out reporter system. Targeting sites within the eGFP ORF in HEK293T cells (stably expressing eGFP) results in loss of eGFP which can be scored by flow

cytometry. We electroporated a plasmid expressing Nme2Cas9 along with sgRNAs targeting 16 N_4CC sites throughout the eGFP ORF. All N_4CC targeted sites showed reduced eGFP expression compared to no sgRNA control, confirming that Nme2Cas9 recognizes a dinucleotide PAM in mammalian cells (Figure 4.3E).

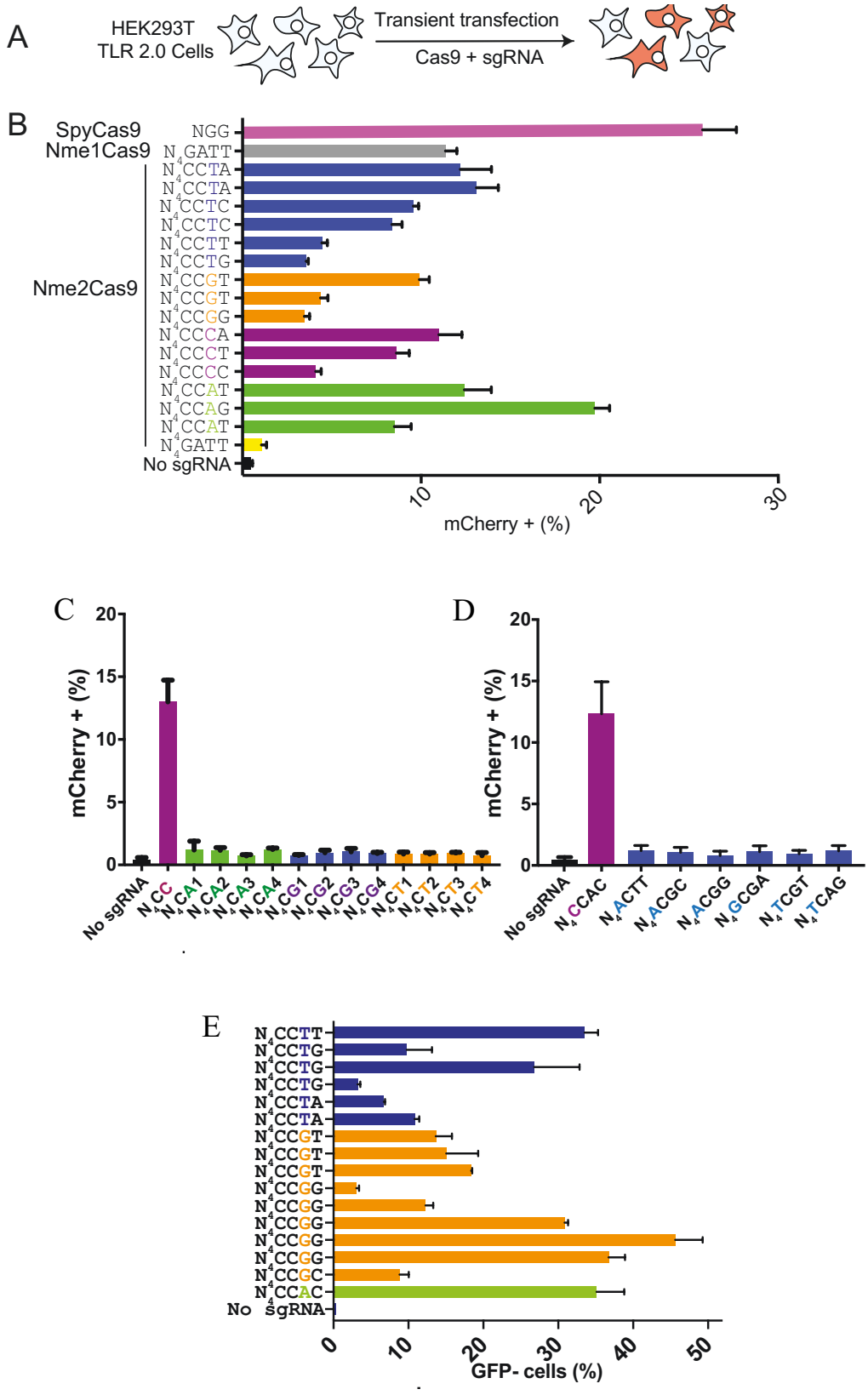
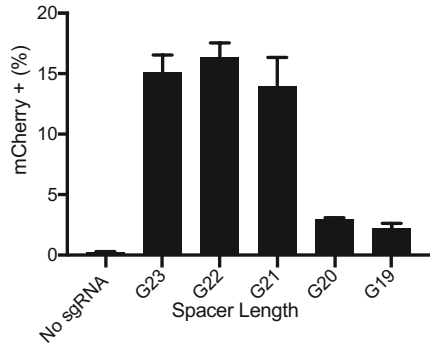
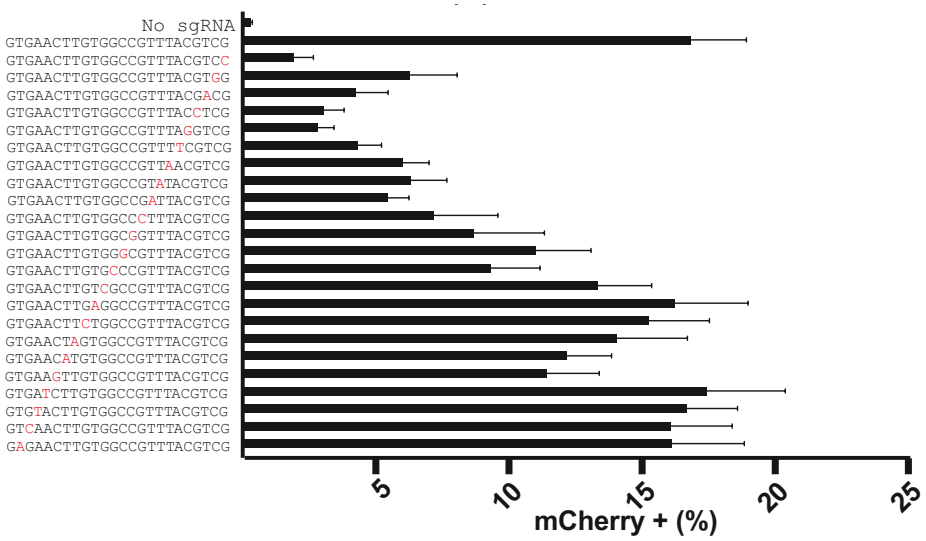

A
HEK293T
TLR 2.0 Cells
Transient transfection
Cas9 + sgRNA
B
SpyCas9
Nme1Cas9
Nme2Cas9
NGG
N4GATT
N4CCTA
N4CCTA
N4CCTC
N4CCTC
N4CCTT
N4CCTG
N4CCGT
N4CCGT
N4CCGG
N4CCCA
N4CCCT
N4CCCC
N4CCAT
N4CCAG
N4CCAT
N4GATT
No sgRNA
10
20
30
mCherry + (%)
C
mCherry + (%)
0
5
10
15
20
No sgRNA
N4CC
N4CA1
N4CA2
N4CA3
N4CA4
N4CG1
N4CG2
N4CG3
N4CG4
N4CT1
N4CT2
N4CT3
N4CT4
D
mCherry + (%)
0
5
10
15
20
No sgRNA
N4CCAC
N4ACTT
N4ACGC
N4ACGG
N4GCGA
N4TCGT
N4TCAG
E
N4CCTT
N4CCTG
N4CCTG
N4CCTG
N4CCTA
N4CCTA
N4CCGT
N4CCGT
N4CCGT
N4CCGG
N4CCGG
N4CCGG
N4CCGG
N4CCGG
N4CCGC
N4CCAC
No sgRNA
0
10
20
30
40
50
GFP- cells (%)

Figure 4.3 Characterization of Nme2Cas9 in mammalian cells
(A) Schematic depicting transient transfection and editing of HEK293T TLR2.0 cells, with mCherry+ cells detected by flow cytometry 72 hours after transfection.
(B) Nme2Cas9 editing of the TLR2.0 reporter. Sites with N_4CC PAMs were targeted with varying efficiencies, while no Nme2Cas9 targeting was observed at an N_4GATT PAM or in the absence of sgRNA. SpyCas9 (targeting a previously validated site with an NGG PAM) and Nme1Cas9 (targeting N_4GATT) were used as positive controls.
(C) Nme2Cas9 targeting at N_4C<u>D</u> sites in TLR2.0, with editing estimated based on mCherry+ cells. Four sites for each non-C nucleotide at the tested position (N_4C<u>A</u>, N_4C<u>T</u> and N_4C<u>G</u>) were examined, and an N_4CC site was used as a positive control.
(D) Nme2Cas9 targeting at N_4<u>D</u>C sites in TLR2.0 [similar to (C)].
(E) Targeting sites within the eGFP ORF in HEK293T cells (stably expressing eGFP) results in loss of eGFP. Electroporation of sgRNAs targeting 16 N_4CC sites throughout the eGFP ORF was used to determine Nme2Cas9 efficiency. All N_4CC targeted sites showed reduced eGFP expression compared to no sgRNA control, confirming that Nme2Cas9 recognizes a dinucleotide PAM in mammalian cells

A


20
15
10
5
0
mCherry + (%)
No sgRNA
G23
G22
G21
G20
G19
Spacer Length

B


No sgRNA
GTGAACTTGTGGCCGTTTACGTCG
GTGAACTTGTGGCCGTTTACGTCC
GTGAACTTGTGGCCGTTTACGTGG
GTGAACTTGTGGCCGTTTACGACG
GTGAACTTGTGGCCGTTTACCTCG
GTGAACTTGTGGCCGTTTAGGTCG
GTGAACTTGTGGCCGTTTTCGTCG
GTGAACTTGTGGCCGTTAACGTCG
GTGAACTTGTGGCCGTATACGTCG
GTGAACTTGTGGCCGATTACGTCG
GTGAACTTGTGGCCCTTTACGTCG
GTGAACTTGTGGCGGTTTACGTCG
GTGAACTTGTGGGCGTTTACGTCG
GTGAACTTGTGCCCGTTTACGTCG
GTGAACTTGTCGCCGTTTACGTCG
GTGAACTTGAGGCCGTTTACGTCG
GTGAACTTCTGGCCGTTTACGTCG
GTGAACTAGTGGCCGTTTACGTCG
GTGAACATGTGGCCGTTTACGTCG
GTGAAGTTGTGGCCGTTTACGTCG
GTGATCTTGTGGCCGTTTACGTCG
GTGTACTTGTGGCCGTTTACGTCG
GTCAACTTGTGGCCGTTTACGTCG
GAGAACTTGTGGCCGTTTACGTCG
5
10
15
20
25
mCherry + (%)

C

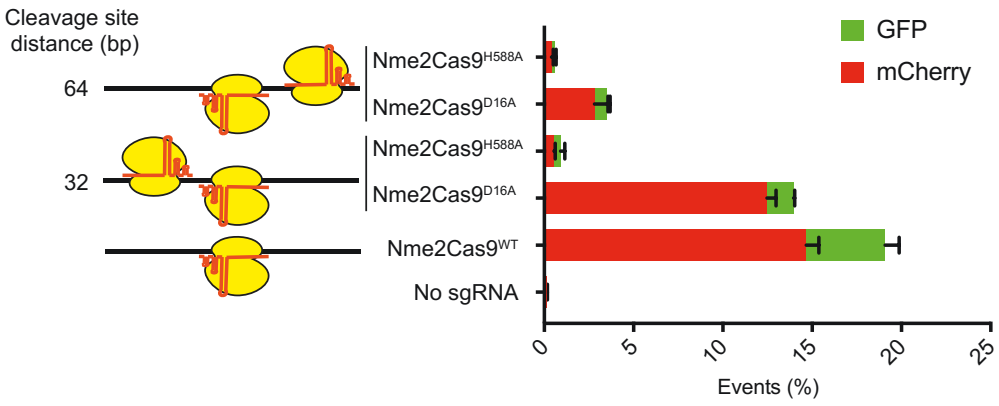

Cleavage site distance (bp)
64
32
Nme2Cas9H588A
Nme2Cas9D16A
Nme2Cas9H588A
Nme2Cas9D16A
Nme2Cas9WT
No sgRNA
GFP
mCherry
0
5
10
15
20
25
Events (%)

Figure 4.4 Further characterization of Nme2Cas9 and its application using TLR
(A) The effect of spacer length on the efficiency of Nme2Cas9 editing. An sgRNA targeting a single TLR2.0 site, with spacer lengths varying from 24 to 20 nts (including the 5'-terminal G required by the U6 promoter), indicate that highest editing efficiencies are obtained with 22-24 nt spacers.
(B) Nme2Cas9 targeting efficiency is differentially sensitive to single-nucleotide mismatches in the seed region of the sgRNA. Data show the effects of walking single-nucleotide sgRNA mismatches along the 23-nt spacer in a TLR2.0 target site.
(C) An Nme2Cas9 dual nickase can be used in tandem to generate NHEJ- and HDR-based edits in TLR2.0. Nme2Cas9- and sgRNA-expressing plasmids, along with an 800-bp dsDNA donor for homologous repair, were electroporated into HEK293T TLR2.0 cells, and both NHEJ (mCherry+) and HDR (GFP+) outcomes were scored by flow cytometry. HNH nickase, Nme2Cas9^{D16A}; RuvC nickase, Nme2Cas9^{H588A}. Cleavage sites 32 bp and 64 bp apart were targeted using either nickase. The HNH nickase (Nme2Cas9^{D16A}) yielded efficient editing, particularly with the cleavage sites that were separated by 32 bp, whereas the RuvC nickase (Nme2Cas9^{H588A}) was not effective. Wildtype Nme2Cas9 was used as a control.

The length of the spacer in the crRNA differs among Cas9 orthologs and can affect on- vs. off-target activity (Cho *et al.*, 2013; Fu *et al.*, 2014). SpyCas9's optimal spacer length is 20 nts, with truncations down to 17 nts tolerated (Fu *et al.*, 2014). In contrast, Nme1Cas9 prefers 24-nt spacers with truncations down to 18-20 nts tolerated (Lee, Cradick and Bao, 2016; Amrani *et al.*, 2018b). To test the spacer length requirements for Nme2Cas9, we created guide RNA plasmids each targeting a single TLR2.0 site, but with varying spacer lengths (Figure 4.4A). We observed comparable activities with G23, G22 and G21 guides, but significantly decreased activity upon further truncation to G20 and G19 lengths (Figure 4.4A). These results validate Nme2Cas9 as a genome editing platform, with 22-24 nt guide sequences, at N_4CC PAM sites in HEK293T cells.

Studies in previously characterized Cas9s have identified a specific region proximal to the PAM where Cas9 activity is highly sensitive to sequence mismatches. This 8 to 12-nt region is known as the seed sequence and has been observed among all Cas9s characterized to date (Gorski, Vogel and Doudna, 2017).To determine whether Nme2Cas9 also possesses a seed sequence, we performed a series of transient transfections each targeting the same locus in TLR2.0, but with a single-nucleotide mismatch at different positions of the guide (Figure 4.4B). We observed a significant decrease in the number of mCherry-positive cells for mismatches in the first 10-12 nts proximal to the PAM, suggesting that Nme2Cas9 possesses a seed sequence in this region.

As described previously, Cas9 enzymes use their HNH and RuvC domains to cleave the guide-complementary and non-complementary strand of the target DNA, respectively. SpyCas9 nickases (nCas9s), in which either the HNH or RuvC domain is mutationally inactivated, have been used to induce homology-directed repair (HDR) and to improve genome editing specificity via DSB induction by dual nickases. To test the efficacy of Nme2Cas9 as a nickase, we created Nme2Cas9D16A (HNH nickase) and Nme2Cas9H588A (RuvC nickase), which possess alanine mutations in catalytic residues of the RuvC and HNH domains, respectively (Esvelt *et al.*, 2013; Zhang *et al.*, 2013). We then used TLR2.0 cells, along with a GFP donor dsDNA, to determine whether Nme2Cas9-induced nicks can induce precise edits via HDR. We used target sites within TLR2.0 to test the functionality of each nickase using guides targeting cleavage sites spaced 32 bp and 64 bp apart (Figure 4.4C). Wildtype Nme2Cas9 targeting a single site showed efficient editing, with both NHEJ and HDR as outcomes of repair. For nickases, cleavage sites 32 bp and 64 bp apart showed editing using the Nme2Cas9D16A (HNH nickase), but neither target pair worked with Nme2Cas9H588A. These results suggest that Nme2Cas9 HNH nickase can be used for efficient genome editing, as long as the sites are in close proximity.

We next tested Nme2Cas9's ability to function in different mammalian cell lines. As an initial test, we targeted 40 different sites (29 with a N_4CC PAM, and 11 sites with a N_4CD PAM) (Appendix 2) throughout the human genome in HEK293T cells using transient transfections. 72-hours post transfection, cells were harvested, followed by genomic DNA extraction and selective amplification of the targeted locus.

We used TIDE to measure indel efficiency at each locus (see Figure 1.5; Brinkman et al, 2014). Nme2Cas9 editing was detectable at most of these sites, even though efficiencies varied depending on the target sequence (Appendix 2). 14 sites with N4CC PAMs were analyzed in triplicate, and consistent editing was observed (Figure 4.5A). In addition, editing efficiency could be improved significantly by increasing the quantity of the Nme2Cas9 plasmid delivered (Figure 4.5B).

We next tested the ability of Nme2Cas9 to function in mouse Hepa1-6 cells (hepatoma-derived). For Hepa1-6 cells, a single plasmid encoding both Nme2Cas9 and an sgRNA (targeting either Rosa26 or Pcsk9) was transiently transfected and indels were measured after 72 hrs. Editing was readily observed at both sites (Figure 4.6A, left). We also tested Nme2Cas9's functionality when stably expressed in human leukemia K562 cells. To this end, we created a lentiviral construct expressing Nme2Cas9 and transduced cells to stably express Nme2Cas9 under the control of the SFFV promoter. This stable cell line did not show any visible differences with respect to growth and morphology in comparison to untransduced cells, suggesting that Nme2Cas9 is not toxic when stably expressed. These cells were transiently electroporated with plasmids expressing sgRNAs and analyzed by TIDE after 72 hours to measure indel efficiencies. We observed efficient (>50%) editing at all three sites tested, validating Nme2Cas9's ability to function upon lentiviral delivery in K562 cells (Figure 4.6A, right).

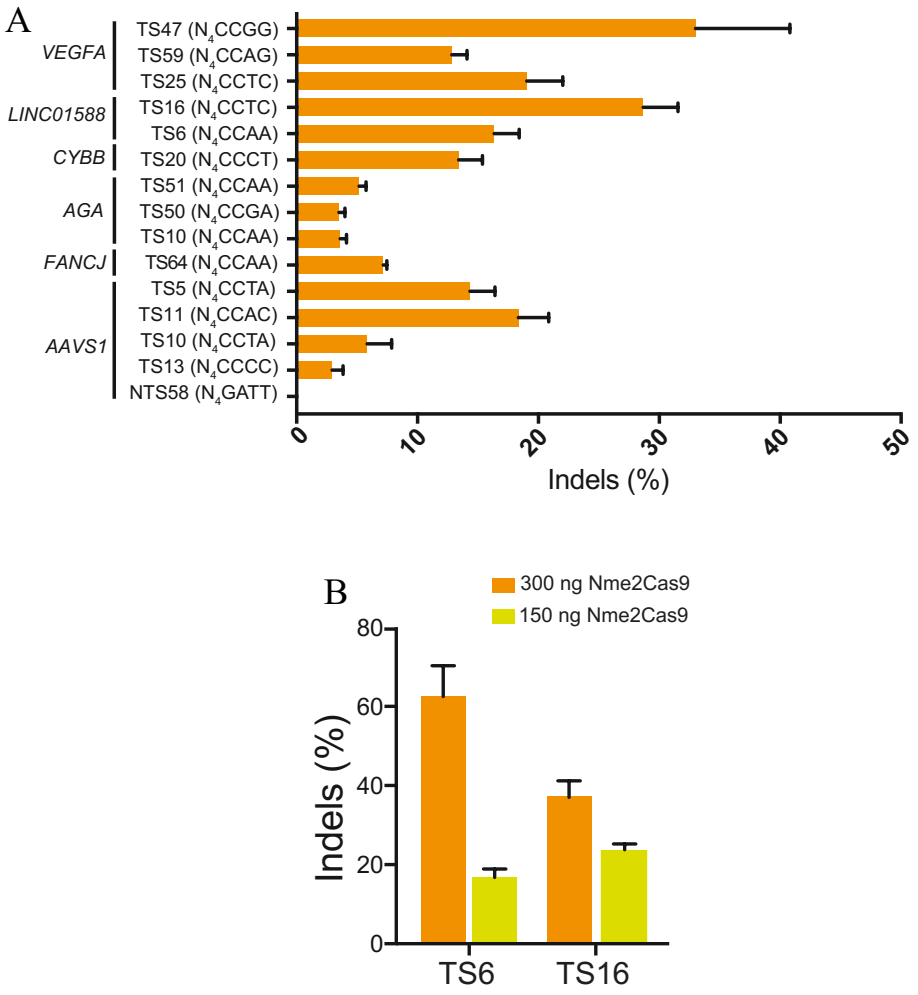


Figure 4.5 Genome editing at endogenous sites in HEK293T cells
(A) Nme2Cas9 genome editing of endogenous human sites in HEK293T cells following transient transfection of Nme2Cas9- and sgRNA-expressing plasmids. 40 sites were screened initially (Table S1); the 14 sites shown (selected to include representatives of varying editing efficiencies, as measured by TIDE) were then re-analyzed in triplicate. An Nme1Cas9 target site (with an N_4GATT PAM) was used as a negative control.
(B) Increasing the dose of electroporated Nme2Cas9 plasmid (500 ng, vs. 200 ng in Figure 3A) improves editing efficiency at two sites (TS16 and TS6). Data provided in yellow are re-used from Figure 3A.

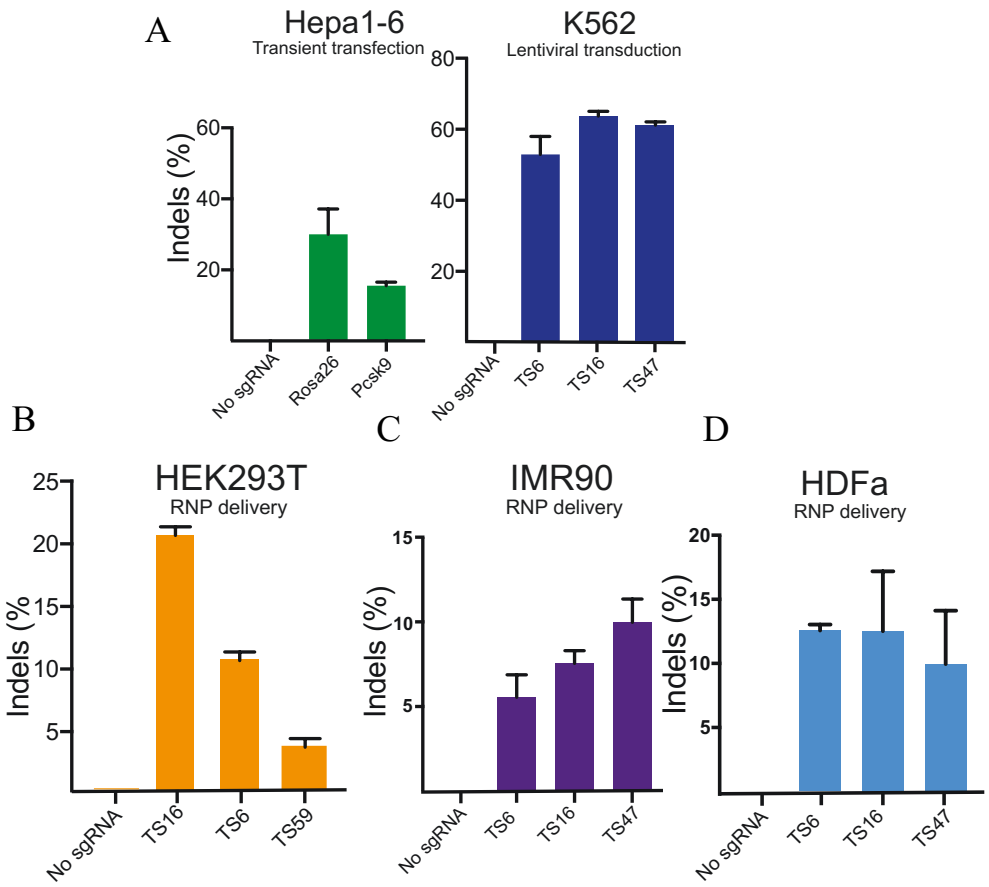


Figure 4.6 Editing by Nme2Cas9 in different cells and via different delivery methods

(A) Left panel: Transient transfection of a single plasmid expressing both Nme2Cas9 and sgRNA (targeting the *Pcsk9* and *Rosa26* loci) enables editing in Hepa1-6 mouse cells, as detected by TIDE. Right panel: Electroporation of sgRNA plasmids into K562 cells stably expressing Nme2Cas9 from a lentivector results in efficient indel formation.

(B, C and D) Nme2Cas9 can be electroporated as an RNP complex to induce genome editing. 40 picomoles Cas9 along with 50 picomoles of sgRNAs targeting three different loci were electroporated into cells. Indels were measured after 72h using TIDE. In (C) 80 picomoles Cas9 along with 100 picomoles were used.

All results represent 3 independent biological replicates, and error bars represent s.e.m.

Ribonucleoprotein (RNP) delivery of Cas9 and its sgRNA is also useful for some genome editing applications, and the greater transience of Cas9's presence can minimize off-target editing (Zuris *et al.*, 2015). Moreover, some cell types (e.g. certain immune cells) are recalcitrant to DNA transfection-based editing (Schumann *et al.*, 2015). To test whether Nme2Cas9 is functional by RNP delivery, we cloned 6xHis-tagged Nme2Cas9 (fused to three NLSs) into a bacterial expression construct and purified the recombinant protein, which we then loaded with T7 RNA polymerase-transcribed sgRNAs targeting three previously validated sites. Electroporation of the Nme2Cas9:sgRNA complex induced successful editing at each of the three target sites in HEK293T cells, as detected by TIDE (Figure 4.6B). To test for editing in cells that are intrinsically unreceptive to DNA, we used RNP electroporations to target several validated sites in primary human dermal fibroblasts adult cells (HDFa) and IMR90 lung fibroblasts. We observed editing at all sites targeted (Figure 4.6C and D), suggesting that Nme2Cas9 RNP complex can be used for editing in primary and difficult-to-transfect cells. Collectively these results indicate that Nme2Cas9 can be delivered effectively via plasmid or lentivirus, or as an RNP complex, in multiple cell types.

4.2.3 Several Acr families can be used as off-switches for Nme2Cas9

To date, five families of Acrs from diverse bacterial species have been shown to inhibit Nme1Cas9 *in vitro* and in human cells (Lee *et al.*, 2018; Pawluk *et al.*, 2016) Considering the high sequence identity between Nme1Cas9 and Nme2Cas9, we reasoned that at least some of these Acr families should inhibit Nme2Cas9. To test this, we expressed and purified all five families of recombinant Acrs, and tested Nme2Cas9's ability to cleave a target *in vitro* in the presence of a member of each family (10:1 Acr:Cas9 molar ratio). We used an inhibitor for the type I-E CRISPR system in E. coli (AcrE2) (Pawluk et al., 2014) as a negative control, while Nme1Cas9 (Pawluk et al., 2016) was used as a positive control. As expected, all 5 families inhibited Nme1Cas9, while AcrE2 failed to do so (Figure 4.6A, top). AcrIIC1$_{Nme}$, AcrIIC2$_{Nme}$, AcrIIC3$_{Nme}$, and AcrIIC4$_{Hpa}$ completely inhibited Nme2Cas9. Strikingly, however, AcrIIC5$_{Smu}$ (the most potent of the Nme1Cas9 inhibitors) did not inhibit Nme2Cas9 *in vitro* even at a 10-fold molar excess, suggesting that it likely inhibits Nme1Cas9 by interacting with its PID (Figure 4.6A, middle). To further test this, we performed the same *in vitro* cleavage assay on the Nme1Cas9/Nme2Cas9 chimera [Nme1Cas9 with the PID of Nme2Cas9 (Figure 4.2B). Due to the reduced activity of this hybrid, we used a ~30x higher concentration of Cas9 to achieve a similar cleavage efficiency while maintaining the 10:1 Acr:Cas9 molar ratio. We observed no inhibition by AcrIIC5Smu on this protein chimera (Figure 4.6A, bottom), providing further evidence that AcrIIC5$_{Smu}$ likely interacts with the PID of Nme1Cas9. Regardless of the

mechanistic basis for the differential inhibition by AcrIIC5$_{Smu}$, our results indicate that Nme2Cas9 is subject to inhibition by the other four type II-C Acr families.

Based on our *in vitro* data, we hypothesized that AcrIIC1$_{Nme}$, AcrIIC2$_{Nme}$, AcrIIC3$_{Nme}$, and AcrIIC4$_{Hpa}$ could be used as off-switches for Nme2Cas9 genome editing. To test this, we performed Nme2Cas9/sgRNA plasmid transfections in HEK293T cells (targeting TS16) in the presence or absence of Acr expression plasmids. We co-transfected 150 ng of each plasmid, as most Acrs inhibited Nme1Cas9 at those plasmid ratios (Pawluk et al., 2016). As expected, AcrIIC1$_{Nme}$, AcrIIC2$_{Nme}$, AcrIIC3$_{Nme}$ and AcrIIC4$_{Hpa}$ inhibited Nme2Cas9 genome editing, while AcrIIC5$_{Smu}$ had no effect (Figure 4.7B). We observed complete inhibition by AcrIIC3$_{Nme}$ and AcrIIC4$_{Hpa}$, suggesting that they have high potency against Nme2Cas9 compared to AcrIIC1$_{Nme}$ and AcrIIC2$_{Nme}$. To further compare the potency of AcrIIC1$_{Nme}$ and AcrIIC4$_{Hpa}$, we repeated the experiments at various ratios of Acr plasmid to Cas9 plasmid (Figure 4.7C). We observe that the AcrIIC4Hpa plasmid is especially potent against Nme2Cas9. Together, these data suggest that several Acr proteins can be used as off-switches for Nme2Cas9-based applications.

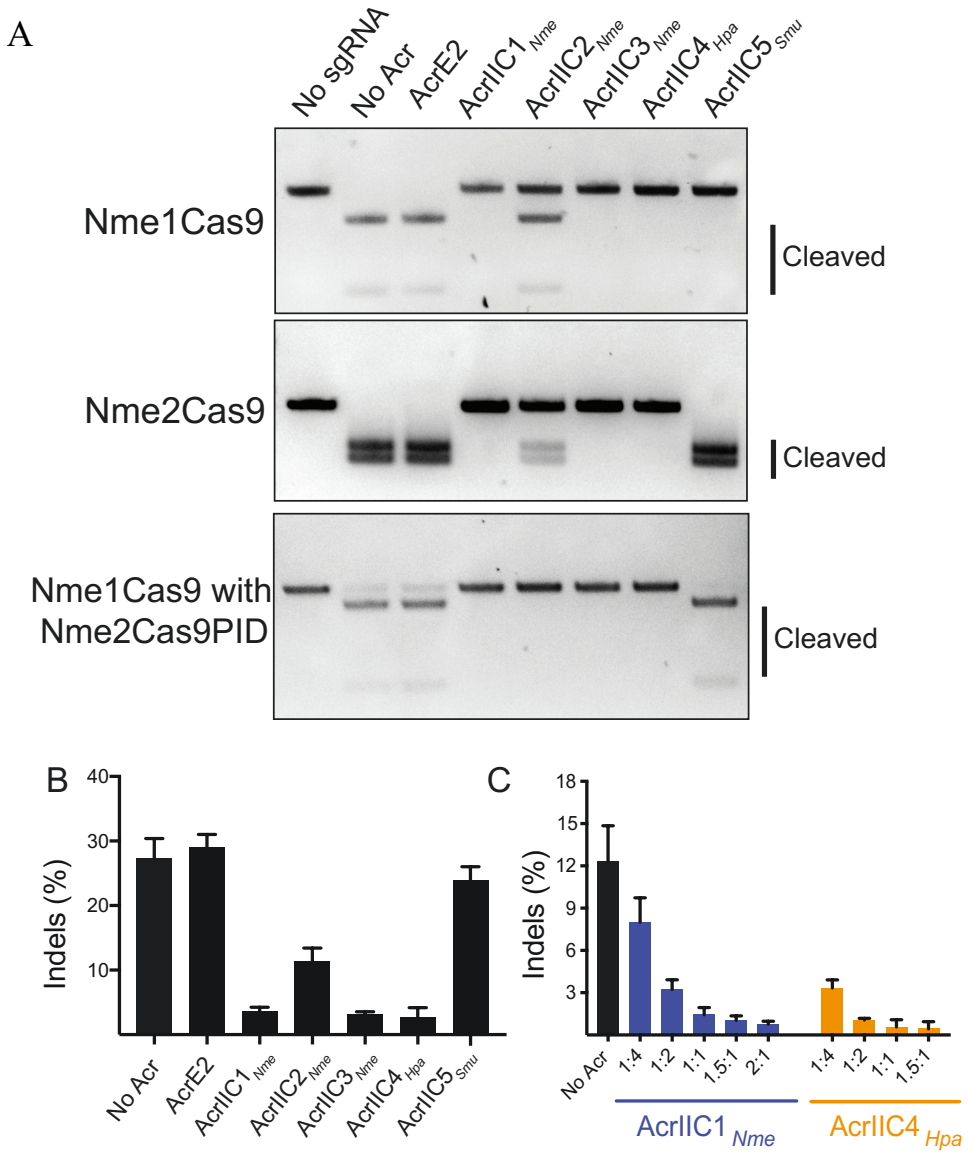


Figure 4.7 Nme2Cas9 is inhibited by a subset of Acrs

(A) *In vitro* cleavage assay of Nme1Cas9 and Nme2Cas9 in the presence of five previously characterized anti-CRISPR proteins (10:1 ratio of Acr:Cas9). Top: All five previously characterized type II-C Acr families inhibited Nme1Cas9, as expected. Middle: Nme2Cas9 inhibition mirrors that of Nme1Cas9, except for the lack of inhibition by AcrIIC5$_{Smu}$. Bottom, *In vitro* cleavage by the Nme1Cas9-Nme2Cas9PID chimera in the presence of previously characterized Acr proteins.

(B) Genome editing in the presence of the five previously described anti-CRISPR families.

(C) Acr inhibition of Nme2Cas9 is dose-dependent with distinct apparent potencies. Nme2Cas9 is fully inhibited by AcrIIC1Nme and AcrIIC4Hpa at 2:1 and 1:1 mass ratios of cotransfected Acr and Nme2Cas9 plasmids, respectively.

4.2.4 Nme2Cas9 is hyper-accurate in mammalian cells

As discussed previously, Nme1Cas9 demonstrates remarkable editing fidelity in cells and mouse models, and the similarity of Nme2Cas9 to Nme1Cas9 over most of its length suggests that it may likewise be hyper-accurate. However, the higher number of sites sampled in the genome as a result of the dinucleotide PAM could create more opportunities for Nme2Cas9 off-targeting in comparison with Nme1Cas9 and its less frequently encountered 4-nucleotide PAM. To assess the off-target profile of Nme2Cas9, we used GUIDE-seq (genome-wide, unbiased identification of double-stranded breaks enabled by sequencing) to identify potential off-target sites empirically, in an unbiased fashion (Tsai *et al.*, 2014). Even the best off-target prediction algorithms are prone to false negatives, necessitating empirical target site profiling methods (Tsai and Joung, 2016). GUIDE-seq relies on the incorporation of double-stranded oligodeoxynucleotides (dsODNs) into DNA double-stranded break sites throughout the genome. These insertion sites are then detected by amplification and high-throughput sequencing.

Because SpyCas9 is by far the best-characterized Cas9 ortholog, it is useful for multiplexed applications with other Cas9s, and as a benchmark for their editing properties (Jiang and Doudna, 2017). We cloned SpyCas9 and Nme2Cas9 into identical plasmid backbones, with the same UTRs, linkers, NLSs, and promoters, for parallel transient transfections (along with similarly matched sgRNA-expressing plasmids) into HEK293T cells. First, we confirmed that the guides for SpyCas9 and Nme2Cas9 are

orthogonal, i.e. that Nme2Cas9 sgRNAs do not direct editing by SpyCas9, and vice versa (Figure 4.8C), as expected based on earlier results with Nme1Cas9 (Fonfara *et al.*, 2014).Next, to enable the use of SpyCas9 as a benchmark for GUIDE-seq, we took advantage of the fact that SpyCas9 and Nme2Cas9 have non-overlapping PAMs and can therefore potentially edit any dual site (DS) flanked by a 5'-NGGNCC-3' sequence, which simultaneously fulfills the PAM requirements of both (Figure 4.8A). This enables side-by-side comparisons of off-targeting with guides that edit the exact same on-target site (Figure 4.8A). We targeted 28 DSs at multiple loci throughout the genome using plasmids expressing each Cas9 and their respective sgRNAs. 72 hours after plasmid delivery, we performed TIDE analysis on the sites targeted by each nuclease. Nme2Cas9 induced indels at 19 sites, albeit at low efficiencies (<5%) at three of them, while SpyCas9 induced indels at 23 of the sites (Figure 4.8B, left). While SpyCas9 is usually more efficient, both enzymes have similar efficiencies at most sites (Figure 4.8B, right).

For GUIDE-seq, we selected DS2, DS4 and DS6 to sample off-target cleavage with Nme2Cas9 guides that direct on-target editing as efficiently, less efficiently, or more efficiently than the corresponding SpyCas9 guides, respectively (Figure 4.9A). In addition to the three dual sites, we added one of the most efficiently edited Nme2Cas9 target sites (TS6, with 30-50% indel efficiency depending on the cell type; see Figure 4.5 and 4.6), along with the mouse Pcsk9 and Rosa26 Nme2Cas9 sites (Figure 4.6A). We performed plasmid transfections of each Cas9 along with their cognate sgRNAs and the dsODNs, and prepared GUIDE-seq libraries. GUIDE-seq analysis (Tsai *et al.*,

2014; Zhu *et al.*, 2017) revealed efficient on-target editing for both Cas9 orthologs (Figure 4.9B), with relative efficiencies (as reflected by GUIDE-seq read counts) that are similar to those observed by TIDE. For off-target identification, the analysis revealed that the DS2, DS4, and DS6 SpyCas9 sgRNAs appeared to direct editing at 93, 10, and 118 candidate off-target sites, respectively, in the normal range of off-targets when plasmid-based SpyCas9 editing is analyzed by GUIDE-seq (Fu *et al.*, 2014). In striking contrast, the DS2, DS4, and DS6 Nme2Cas9 sgRNAs appeared to direct editing at 1, 0, and 1 off-target sites, respectively (Figure 4.9C). When compared to the GUIDE-seq read counts for the SpyCas9 off-targets (Figure 4.9D), those of Nme2Cas9 were very low (Figure 4.9E), further suggesting that Nme2Cas9 is highly specific. Nme2Cas9 GUIDE-seq analyses with the TS6, Pcsk9, and Rosa26 yielded similar results (0, 0, and 1 off-target sites, respectively, with a modest read count for the Rosa26-OT1 off-target site) (Figure 4.9C and E).

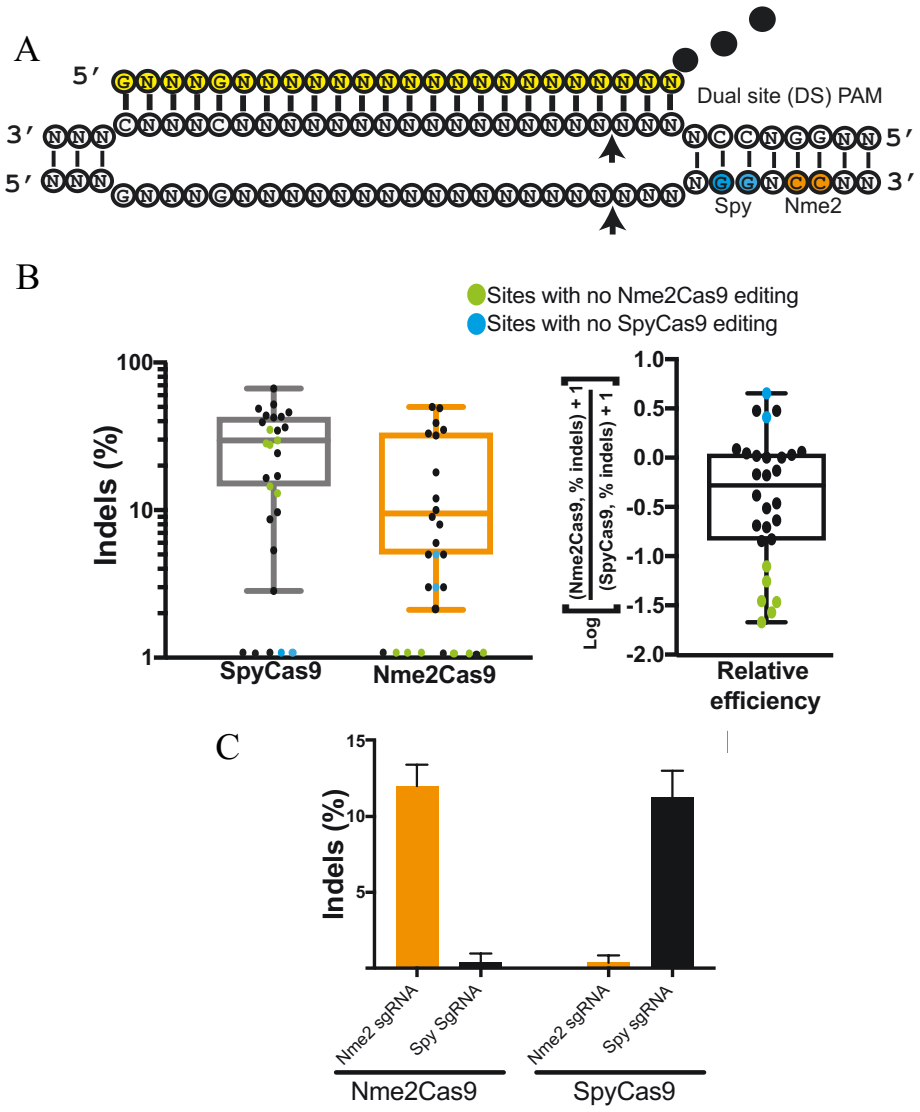


Figure 4.8 Comparing the efficiencies of Nme2Cas9 and SpyCas9
(A) Schematic depicting dual sites (DSs) targetable by both SpyCas9 and Nme2Cas9
(B) Nme2Cas9 and SpyCas9 efficiencies vary based on the locus and target site. Box-and-whisker plots indicate editing efficiencies at 28 dual sites by Nme2Cas9 and SpyCas9 (left). The sites that showed no editing were excluded from the analysis. Relative efficiencies of Nme2Cas9 and SpyCas9 show that Nme2Cas9 is, on average, less efficient than SpyCas9 (right). Editing efficiencies by both Cas9 orthologs at all 28 sites were included in the analysis of relative efficiencies on the right.
(C) Nme2Cas9 and SpyCas9 guides are orthogonal.

To validate the off-target sites detected by GUIDE-seq, we performed targeted deep sequencing to measure indel formation at the top off-target loci following GUIDE-seq-independent editing (i.e. without co-transfection of the dsODN). While SpyCas9 showed considerable editing at most off-target sites tested (in some instances, more efficient than that at the corresponding on-target site), Nme2Cas9 exhibited no detectable indels at the lone DS2 and DS6 candidate off-target sites (Figure 4.10A). With the Rosa26 sgRNA, Nme2Cas9 induced ~1% editing at the Rosa26-OT1 site in Hepa1-6 cells, compared to ~30% on-target editing (Figure 4.10A). It is noteworthy that this off-target site has a consensus Nme2Cas9 PAM (ACTCCCT) with only 3 mismatches at the PAM-distal end of the guide-complementary region (i.e. outside of the seed) (Figure 4.10B). These data support and reinforce our GUIDE-seq results indicating a high degree of accuracy for Nme2Cas9 genome editing in mammalian cells.

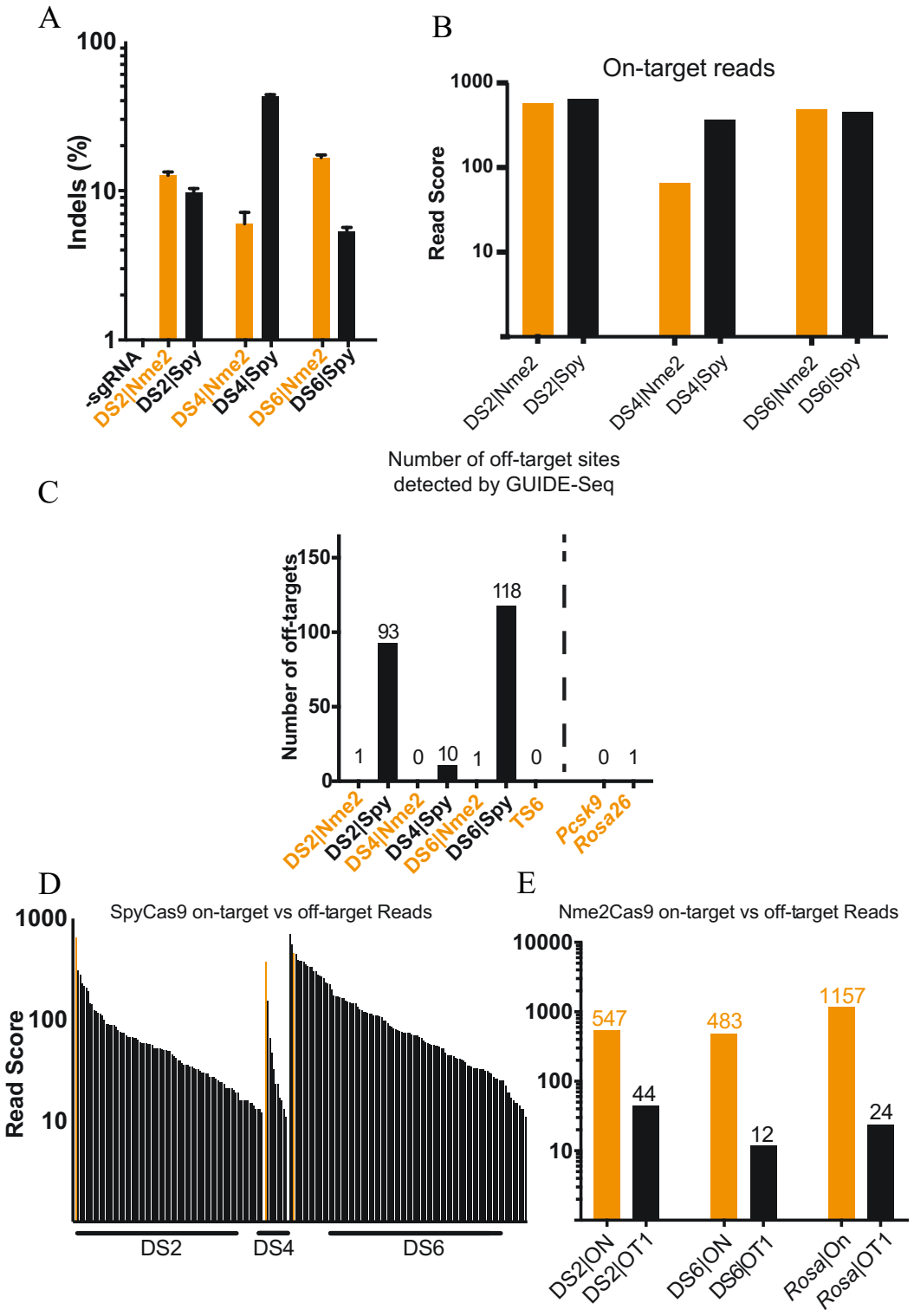

A
Indels (%)
100
10
1
-sgRNA
DS2|Nme2
DS2|Spy
DS4|Nme2
DS4|Spy
DS6|Nme2
DS6|Spy
B
On-target reads
Read Score
1000
100
10
DS2|Nme2
DS2|Spy
DS4|Nme2
DS4|Spy
DS6|Nme2
DS6|Spy
C
Number of off-target sites
detected by GUIDE-Seq
Number of off-targets
150
100
50
0
1
93
0
10
1
118
0
0
1
DS2|Nme2
DS2|Spy
DS4|Nme2
DS4|Spy
DS6|Nme2
DS6|Spy
TS6
Pcsk9
Rosa26
D
SpyCas9 on-target vs off-target Reads
Read Score
1000
100
10
DS2
DS4
DS6
E
Nme2Cas9 on-target vs off-target Reads
10000
1000
100
10
547
44
483
12
1157
24
DS2|ON
DS2|OT1
DS6|ON
DS6|OT1
Rosa|On
Rosa|OT1

Figure 4.9 Nme2Cas9 is highly accurate in human cells
(A) DS2, DS4 and DS6 were selected for GUIDE-Seq analysis as Nme2Cas9 was equally efficient, less efficient and more efficient than SpyCas9, respectively, at these sites.
(B) Nme2Cas9 and SpyCas9 exhibit comparable on-target editing efficiencies as assessed by GUIDE-seq. Bars indicate on-target read counts from GUIDE-Seq at the three dual sites targeted by each ortholog. Orange bars represent Nme2Cas9 and black bars represent SpyCas9.
(C) Numbers of off-target sites detected by GUIDE-Seq for each nuclease at individual target sites are shown. In addition to dual sites, we analyzed TS6 (because of its high on-target editing efficiency) and *Pcsk9* and *Rosa26* sites in mouse Hepa1-6 cells (to measure accuracy in another cell type).
(C) SpyCas9's on-target vs. off-target read counts for each site. Orange bars represent the on-target reads while black bars represent off-targets.
(D) Nme2Cas9's on-target vs. off-target reads for each site.

To further corroborate our GUIDE-Seq results, we used CRISPRseek (Zhu *et al.*, 2014) to computationally predict potential off-target sites for two of our most active Nme2Cas9 sgRNAs (targeting TS25 and TS47, both of which are also in VEGFA). We selected three (TS25) or four (TS47) of the most closely matched predicted sites, five with N_4CC PAMs and two with N4CA PAMs; each had 2-5 mismatches, mostly in their PAM-distal, non-seed regions (Figure 4.10C). We then compared on- vs. off-target editing (after Nme2Cas9+sgRNA plasmid transfections into HEK293T cells) by targeted amplification of each locus, followed by TIDE analysis. Consistently, no indels could be detected at those off-target sites for either sgRNA by TIDE, while efficient on-target editing was readily detected in DNA from the same populations of cells. Taken together, our data indicate that Nme2Cas9 is a naturally hyper-accurate genome editing platform in mammalian cells.

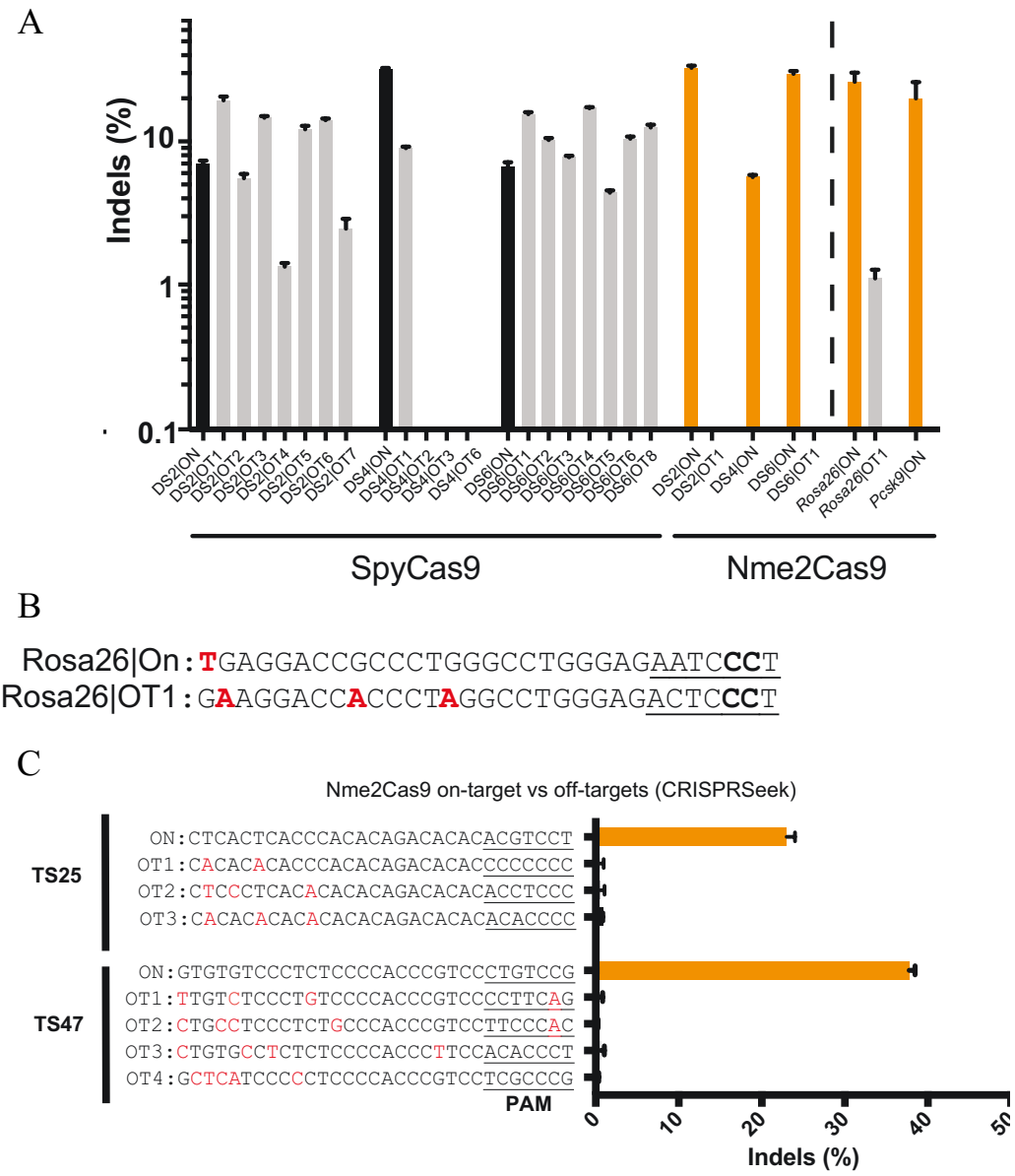


Figure 4.10 Deep-sequencing and CRISPRSeek confirm Nme2Cas9's accuracy
(A) Targeted deep sequencing to detect indels in edited cells confirms the high Nme2Cas9 accuracy indicated by GUIDE-seq.
(B) Sequence for the validated off-target site of the *Rosa26* guide, showing the PAM region (underlined), the consensus CC PAM dinucleotide (bold), and three mismatches in the PAM-distal portion of the spacer (red).
(C) Bar graphs showing indel efficiencies (measured by TIDE) at potential off-target sites predicted by CRISPRSeek. On- and off-target site sequences are shown on the left, with the PAM region underlined and sgRNA mismatches and non-consensus PAM nucleotides given in red.

4.2.5 Nme2Cas9 is functional *in vivo* and ex vivo by all-in-one AAV delivery

The compact size, small PAM, and high fidelity of Nme2Cas9 offer major advantages for *in vivo* genome editing using AAV. To test whether effective Nme2Cas9 genome editing can be achieved via single-AAV delivery, we cloned Nme2Cas9 with its sgRNA and their promoters (U1a and U6, respectively) into an AAV vector backbone (4.11A). We then packaged this all-in-one AAV.sgRNA.Nme2Cas9 into a hepatotropic AAV8 capsid to target two genes in the mouse liver: Rosa26 (a commonly used safe harbor locus for transgene insertion) as a negative control (Friedrich and Soriano, 1991), and Pcsk9, a major regulator of circulating cholesterol homeostasis as a phenotypic target (Rashid *et al.*, 2005). SauCas9- or Nme1Cas9-induced indels in Pcsk9 in the mouse liver results in reduced cholesterol levels, providing a useful and easy-to-score *in vivo* benchmark for new editing platforms (Ran *et al.*, 2015; Ibraheim *et al.*, 2018b). The Nme2Cas9 guides were the same as those used above (Figure 4.6A); because Rosa26-OT1 was the only Nme2Cas9 off-target site that we were able to validate in cultured mammalian cells (Figure 4.10A and B), the Rosa26 guide also provided us with an opportunity to assess on- vs. off-target editing *in vivo*.

The tail veins of two groups of mice (n = 5) were injected with 4×10^{11} AAV8.sgRNA.Nme2Cas9 genome copies (GCs) targeting either Pcsk9 or Rosa26. Serum was collected at 0, 14 and 28 days post-injection for cholesterol level measurement. Mice were sacrificed at 28 days post-injection and liver tissues were harvested (Figure 4.11A). Targeted deep sequencing of each locus revealed ~38% and

~46% indel induction at the Pcsk9 and Rosa26 editing sites, respectively, in the liver (Figure 4.11B). Because hepatocytes constitute only 65-70% of total cellular content in the adult liver (Racanelli and Rehermann, 2006), Nme2Cas9 AAV-induced hepatocyte editing efficiencies with sgPcsk9 and sgRosa were approximately 54-58% and 66-71%, respectively. We detected only 2.25% liver indels overall (~3-3.5% in hepatocytes) at the Rosa26-OT1 off-target site (Figure 4.11B), comparable to the 1% editing that we observed at this site in transfected Hepa1-6 cells (Figure 4.6A). At both 14 and 28 days post-injection, Pcsk9 editing was accompanied by a ~44% reduction in serum cholesterol levels, whereas mice treated with the sgRosa26-expressing AAV maintained normal level of cholesterol throughout the study (Figure 4.11C). The ~44% reduction in serum cholesterol in the Nme2Cas9/sgPcsk9 AAV-treated mice compares well with the ~40% reduction reported with SauCas9 all-in-one AAV when targeting the same gene (Ran et al., 2015). We next performed Western blotting using an anti-PCSK9 antibody to estimate PCSK9 protein levels in the livers of mice treated with sgPcsk9 and sgRosa26. Liver PCSK9 was below the detection limit in mice treated with sgPcsk9, whereas sgRosa26-treated mice exhibited normal levels of PCSK9 (Figure 4.11D). Finally, hematoxylin and eosin (H&E) staining and histology revealed no signs of toxicity or tissue damage in either group after Nme2Cas9 expression (Figure 4.11E). These data validate Nme2Cas9 as a highly effective genome editing system *in vivo*, including when delivered by single -AAV vectors.

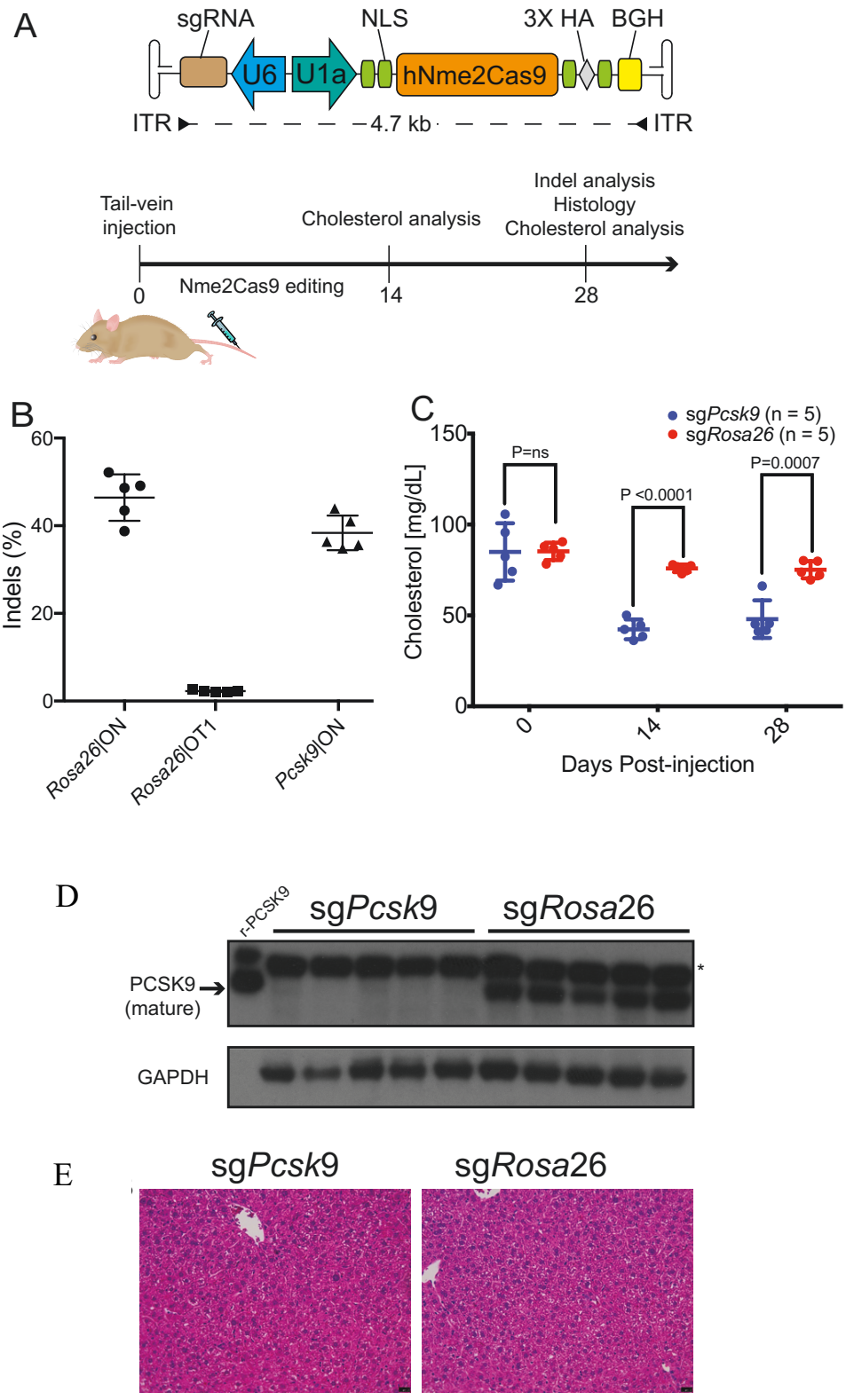

A
sgRNA
NLS
3X HA
BGH
U6
U1a
hNme2Cas9
ITR
4.7 kb
ITR
Tail-vein injection
Cholesterol analysis
Indel analysis
Histology
Cholesterol analysis
0
Nme2Cas9 editing
14
28
B
Indels (%)
60
40
20
0
Rosa26|ON
Rosa26|OT1
Pcsk9|ON
C
sgPcsk9 (n = 5)
sgRosa26 (n = 5)
Cholesterol [mg/dL]
150
100
50
0
P=ns
P <0.0001
P=0.0007
0
14
28
Days Post-injection
D
r-PCSK9
sgPcsk9
sgRosa26
PCSK9
(mature)
*
GAPDH
E
sgPcsk9
sgRosa26

Figure 4.11 All-in-one AAV delivery of Nme2Cas9 *in vivo*
(A) Workflow for delivery of AAV8.sgRNA.Nme2Cas9 to lower cholesterol levels in mice by targeting *Pcsk9*. Top: schematic of the all-in-one AAV vector expressing Nme2Cas9 and the sgRNA (individual genome elements not to scale). BGH, bovine growth hormone poly(A) site; HA, epitope tag; NLS, nuclear localization sequence; h, human-codon-optimized. Bottom: Timeline for AAV8.sgRNA.Nme2Cas9 tail-vein injections (4×10^{11} GCs), followed by cholesterol measurements at day 14 and indel, histology and cholesterol analyses at day 28 post-injection.
(B) TIDE analysis to measure indels in DNA extracted from livers of mice injected with AAV8.Nme2Cas9+sgRNA targeting *Pcsk9* and *Rosa26* (control) loci. Indel efficiencies at the lone off-target site identified by GUIDE-seq for these two sgRNAs (Rosa26|OT1) were also assessed by TIDE.
(C) Reduced serum cholesterol levels in mice injected with the *Pcsk9*-targeting guide compared to the *Rosa26*-targeting controls. *P* values are calculated by unpaired two-tailed t-test.
(D) Western blotting using anti-PCSK9 antibody reveals strongly reduced levels of PCSK9 in the livers of mice treated with sg*Pcsk9*, compared to mice treated with sg*Rosa26.* 2ng of recombinant PCSK9 was used as a mobility standard (left-most lane), and a cross-reacting band in the liver samples is indicated by an asterisk. GAPDH was used as loading control (bottom panel).
(E) H&E staining from livers of mice injected with AAV8.Nme2Cas9+sg*Rosa26* (left) or AAV8.Nme2Cas9+sg*Pcsk9* (right) vectors. Scale bars, 25 um.

AAV vectors have recently been used for the generation of genome-edited mice, without the need for microinjection or electroporation, simply by soaking the zygotes in culture medium containing AAV vector(s), followed by reimplantation into pseudopregnant females (Yoon *et al.*, 2018). Editing was obtained previously with a dual-AAV system in which SpyCas9 and its sgRNA were delivered in separate vectors (Yoon et al., 2018). To test whether Nme2Cas9 could enable accurate and efficient editing in mouse zygotes with an all-in-one AAV delivery system, we targeted Tyrosinase (Tyr), bi-allelic inactivation of which disrupts melanin production, resulting in albino pups (Yokoyama *et al.*, 1990). We first validated an efficient Tyr sgRNA (which cleaves the Tyr locus only 17 bp from the site of the classic albino mutation) in Hepa1-6 cells by transient transfections (Figure 4.12A). Next, we incubated C57BL/6NJ zygotes for 5-6 hours in culture medium containing $3x10^9$ or $3x10^8$ GCs of an all-in-one AAV6 vector expressing Nme2Cas9 along with the Tyr sgRNA. After overnight culture in fresh media, those zygotes that advanced to the two-cell stage were transferred to the oviduct of pseudopregnant recipients and allowed to develop to term (Figure 4.12B). Coat color analysis of pups revealed mice that were albino, chinchilla (indicating a hypomorphic allele of Tyrosinase), or that had variegated coat color composed of albino and chinchilla spots but lacking black pigmentation (Figure 4.12C). These results suggest a high frequency of biallelic mutations since the presence of a wild-type Tyrosinase allele should render black pigmentation. A total of five pups (10%) were born from the 3x109 GCs experiment. All of them carried indels; phenotypically, two were albino, one was chinchilla, and two had variegated

pigmentation, indicating mosaicism. From the $3x10^8$ GCs experiment, we obtained 4 pups (14%), two of which died at birth, preventing coat color or genome analysis. Coat color analysis of the remaining two pups revealed one chinchilla and one mosaic pup. These results indicate that single-AAV delivery of Nme2Cas9 and its guide can be used to generate mutations in mouse zygotes without microinjection or electroporation. To measure on-target indel formation in the Tyr gene, we isolated DNA from the tails of each mouse, amplified the locus, and performed TIDE analysis. We found that all mice had high levels of on-target editing by Nme2Cas9, varying from 84% to 100% (Table 4.1). Most lesions in albino mouse 9-1 were either a 1- or a 4-bp deletion (Figure 4.12D), suggesting either mosaicism or trans-heterozygosity. Albino mouse 9-2 exhibited a uniform 2-bp deletion (Figure 4.12D). We cannot conclude that there was no mosaicism in mouse 9-2, or that additional alleles were absent from mouse 9-1, because only tail samples were sequenced and other tissues could have distinct lesions. Analysis of tail DNA from chinchilla mice revealed the presence of in-frame mutations that are potentially the cause of the chinchilla coat color. The limited mutational complexity suggests that editing occurred early during embryonic development in these mice. These results provide a streamlined route toward mammalian mutagenesis through the application of a single AAV vector, in this case delivering both Nme2Cas9 and its sgRNA.

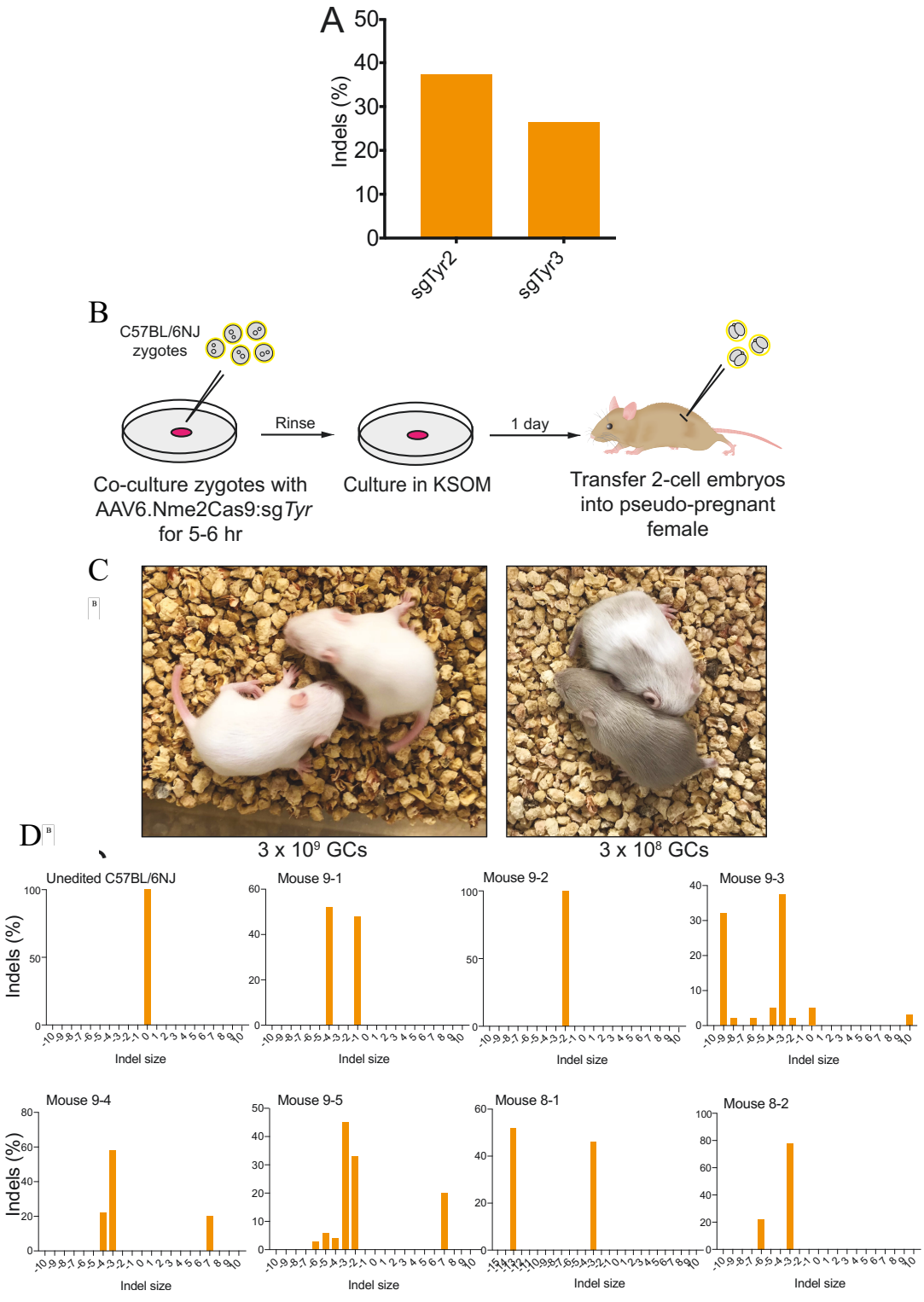

A
Indels (%)
50
40
30
20
10
0
sgTyr2
sgTyr3
B
C57BL/6NJ
zygotes
Rinse
1 day
Co-culture zygotes with
AAV6.Nme2Cas9:sgTyr
for 5-6 hr
Culture in KSOM
Transfer 2-cell embryos
into pseudo-pregnant
female
C
3 x 10^9 GCs
3 x 10^8 GCs
D
Unedited C57BL/6NJ
Mouse 9-1
Mouse 9-2
Mouse 9-3
Mouse 9-4
Mouse 9-5
Mouse 8-1
Mouse 8-2
Indels (%)
Indel size

Figure 4.12 *ex vivo* genome editing with Nme2Cas9
(A) Two sites in *Tyr*, each with N_4CC PAMs, were tested for editing in Hepa1-6 cells. The sg*Tyr2* guide exhibited higher editing efficiency and was selected for further testing. (B) Seven mice survived post-natal development, and each exhibited coat color phenotypes as well as on-target editing, as assayed by TIDE.
(B) Albino (left) and light gray or variegated (middle) mice generated by treatment of zygotes with 3 × 109 GCs, and light gray or variegated mice (right) with 3 × 108 GCs, of AAV6.Nme2Cas9:sgTyr.
(C) Indel spectra from tail DNA of each edited mouse as well as an unedited C57BL/6NJ mouse, as indicated by TIDE analysis. Efficiencies of insertions (positive) and deletions (negative) of various sizes are indicated.

A

AAV Dosage	Number of embryos transferred	Number of pups recovered	Number of *Tyr*-edited pups (albino)	*Tyr* editing frequency (%)
3 x 10^9 GCs	50	5	5 (2)	100
3 x 10^8 GCs	28	4	2(0)	100

B

Mouse ID	AAV dosage	Indels (%)	Coat color
9-1	3 x 10^9 GCs	100	Albino
9-2	3 x 10^9 GCs	100	Albino
9-3	3 x 10^9 GCs	84	mosaic and hypomorphic
9-4	3 x 10^9 GCs	98	mosaic and hypomorphic
9-5	3 x 10^9 GCs	100	mosaic and hypomorphic
8-1	3 x 10^8 GCs	100	mosaic and hypomorphic
8-2	3 x 10^8 GCs	96	mosaic and hypomorphic

Table 4.1 Outcomes of ex vivo Nme2Cas9 delivery in mouse zygotes
(A) Summary of Nme2Cas9.sg*Tyr* single-AAV ex vivo *Tyr* editing experiments at two AAV doses.
(B) Seven mice survived post-natal development, and each exhibited coat color phenotypes as well as on-target editing, as assayed by TIDE.

4.2.6 The structure of Nme2Cas9 reveals the mechanism of PAM recognition

The crystal structure of Nme1Cas9 ortholog was resolved by Yanli Wang's group. Although the structures and interactions of the N-terminal 820 amino acids of Nme2Cas9 should be highly similar to those of Nme1Cas9 (due to their >98% identity in these regions), the PID of the two orthologs are far more divergent, resulting in their distinct PAM specificities. To investigate the molecular mechanism of Nme2Cas9 PAM recognition, in collaboration with Yanli Wang's lab, we determined the crystal structure of Nme2Cas9 in complex with a sgRNA, together with a partially duplexed target DNA containing a N_4CC PAM sequence (Figure 4.13A). dNme2Cas9 was used to prevent DNA cleavage. We solved the co-crystal structures with two distinct dsDNAs with a complete protospacer in the TS, but with either no protospacer or a 5-nucleotide protospacer in the NTS. These structures were solved at 3.2 Å and 2.93 Å resolution, respectively. In both DNA-bound Nme2Cas9 complexes, the HNH domain is completely disordered (Figure 4.13A). Given that these two structures are similar overall, and that the latter has higher resolution and better electron density, our discussion is limited to the structure of Nme2Cas9-sgRNA in complex with DNA containing a 5-nucleotide protospacer in the NTS. Only two NTS protospacer nucleotides are observed, while the other three are disordered. The dsDNA is positioned in the cleft formed by the PID. In the PAM region, the side-chain of Asp1028 accepts a hydrogen bond from the exocyclic amino group of C(-5)' of the NTS (Figure 4.13B). T or G residues in this position would present a keto oxygen that cannot act a hydrogen bond donor. In addition, if adenine were to replace cytosine, the distance

between the side-chain of Asp1028 and the exocyclic -NH_2 group would be too great to form a favorable hydrogen bond. These observations therefore explain the strong requirement for a C at that position, seen in Figure 4.3 In contrast to NTS recognition of C(-5)', the side-chain of Arg1033 forms a pair of hydrogen bonds with the base of G(-6) on the TS, indicating that Nme2Cas9 (like Nme1Cas9) uses both the TS and the NTS for PAM recognition. The other nucleotides within the duplex dsDNA have no base contacts with Nme2Cas9.

To test the importance of the apparent Nme2Cas9 PAM-interacting residues Asp1028 and Arg1033, as well as a neighboring residue (Asn1031) as a negative control, we mutated each to alanine and examined editing efficiencies in transientlytransfected HEK293Tcells. The Asp1028Ala and Arg1033Ala mutations completely abolished Nme2Cas9 editing, whereas the Asn1031Ala mutation had no effect (Figures 4.13C). These results confirm the importance of the crystallographically implicated Asp1028 and Arg1033 interactions in PAM recognition and reveal the basis for differential PAM recognition by Nme1Cas9 and Nme2Cas9

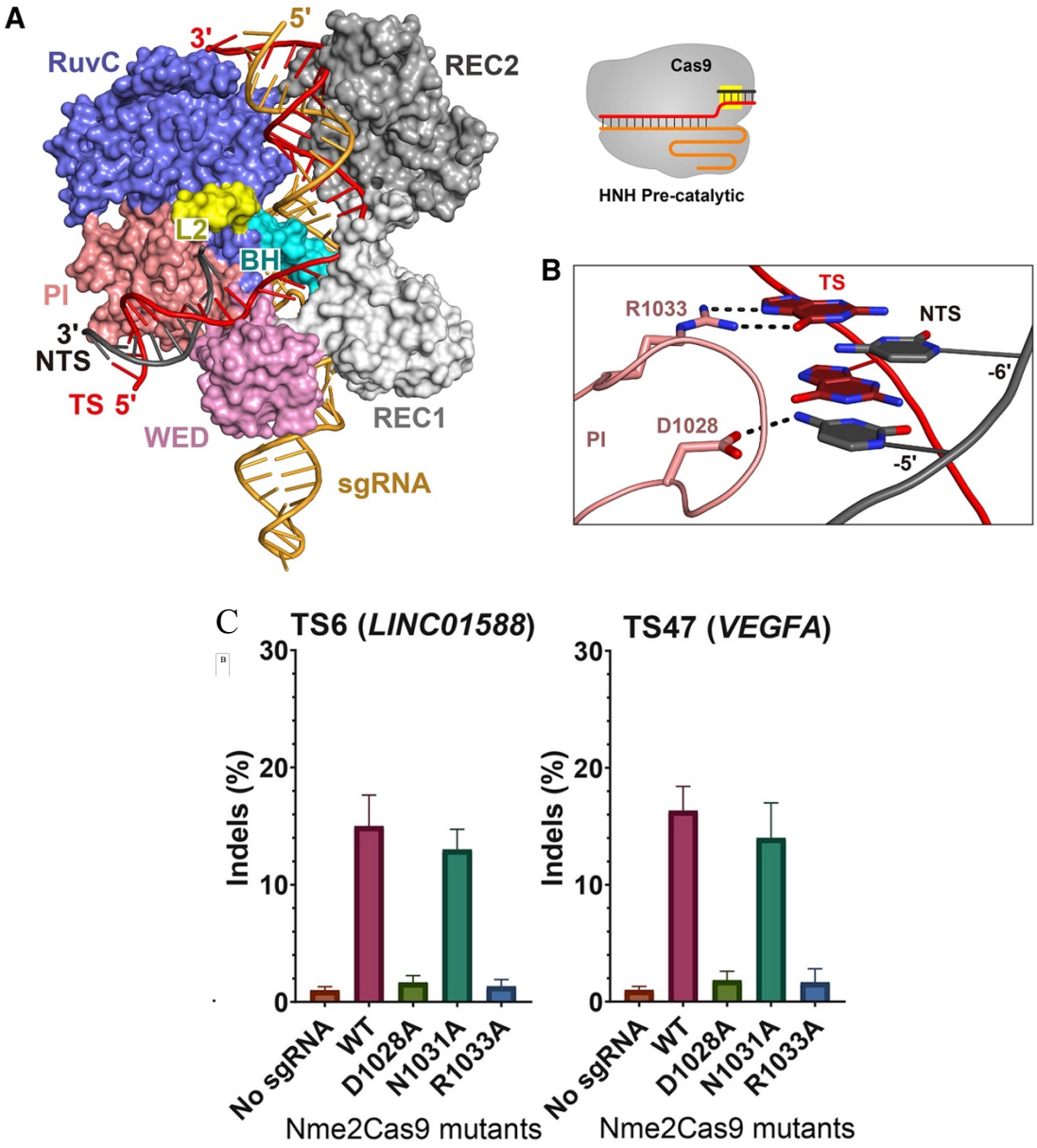


Figure 4.13 The structure of Nme2Cas9 reveals the mechanism of PAM recognition

(A) The overall structure of the Nme2Cas9-sgRNA-dsDNA complex in a pre-catalytic state.

(B) Detailed view of the interaction between the PI domain of Nme2Cas9 and the 5′-N4CC-3′ PAM sequence.

(C) Editing efficiency of wild-type and mutant Nme2Cas9 at 2 genomic sites in HEK293T cells.

4.2.7 Rapidly evolving PIDs are not limited to *N. meningitidis*

Finally, we wondered if the rapidly evolving PIDs could be found in other bacterial clades from which Cas9s for genome editing had been derived. Haemophilus species are in the Gammaproteobacteria class of bacteria and can cause respiratory tract infections. Many Haemophilus strains are sequenced, and since we already established HpaCas9 from *Haemophilus parainfluenzae* for genome editing, we focused on other closely related Haemophilus species. We found Cas9s from several *Haemophilus influenzae* (HinCas9) species were 93% identical to that of HpaCas9, with PIDs that were only 77% identical, suggesting that these Cas9s may have evolved different PAM recognitions.

To investigate whether the HinCas9 ortholog recognized a distinct PAM, we replaced the PID of HpaCas9 with that of HinCas9 and used these recombinant chimeras for *in vitro* PAM identification similar to 3. In contrast to HpaCas9's N_4GATTT, HinCas9 recognizes an N_4GAAA PAM, which corroborates our hypothesis that the high diversity found in the PID resulted in unique PAM recognition. These results suggest that divergent PIDs are likely to be widespread in bacterial species.

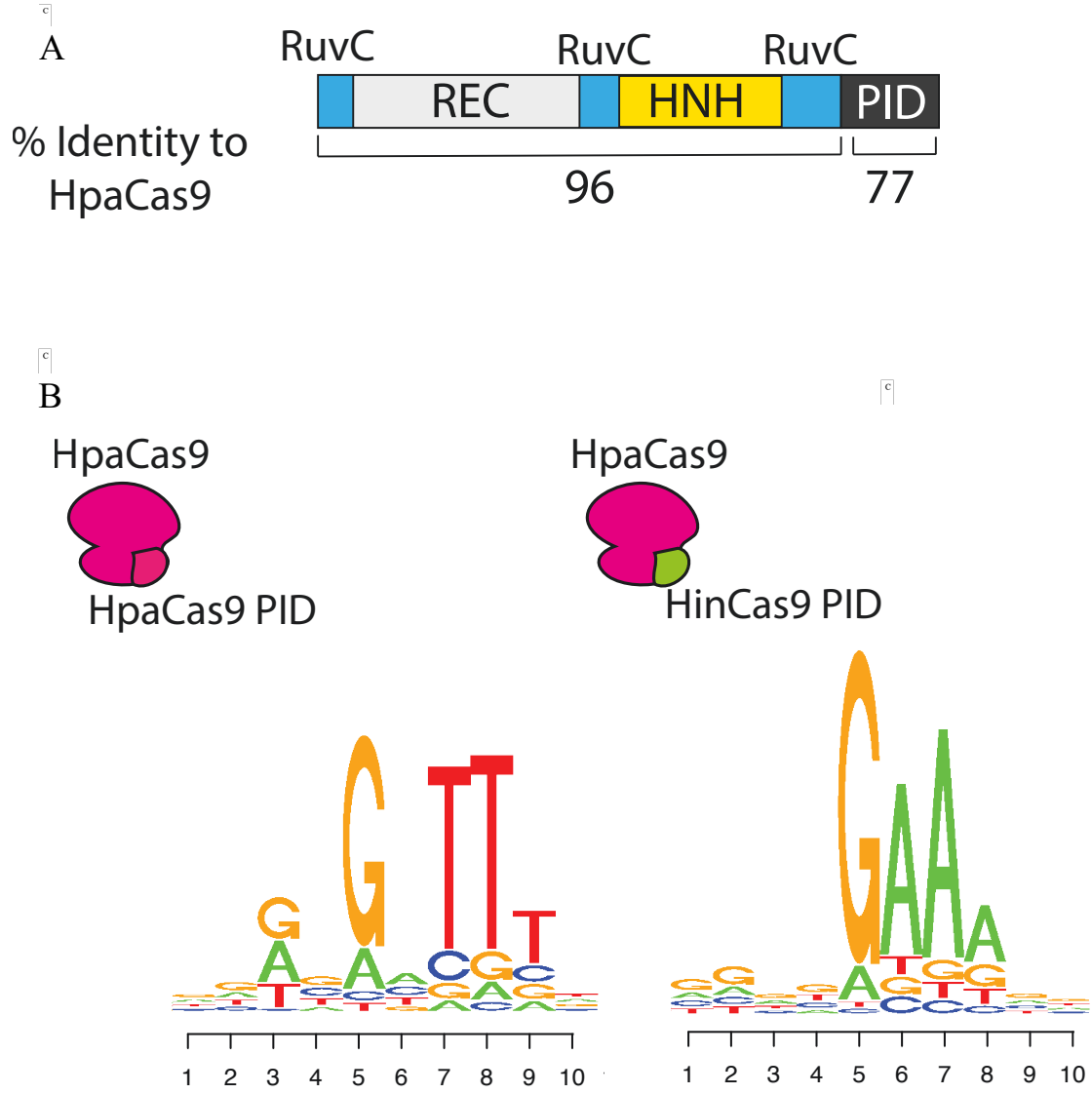


Figure 4.14 Rapidly evolving PIDs found in *Haemophilus* species
(A) Domain architure of HinCas9 and its homology with HpaCas9. Notice the highly divergent C-terminal PID
(B) *In vitro* PAM discovery assay on HpaCas9 (left) and HinCas9's PID swapped onto HpaCas9 (right)

4.3 DISCUSSION

Although compact Cas9 orthologs have been previously validated for genome editing, including via single-AAV delivery, their longer PAMs have restricted therapeutic development due to target site frequencies that are lower than that of the more widely adopted SpyCas9. In addition, SauCas9 and its KKH variant with relaxed PAM requirements are prone to off-target editing with some sgRNAs (Kleinstiver *et al.*, 2015). These limitations are exacerbated with target loci that require editing within a narrow sequence window, or that require precise segmental deletion. We have identified Nme2Cas9 as a compact and highly accurate Cas9 with a less restrictive dinucleotide PAM for genome editing by AAV delivery *in vivo*. The development of Nme2Cas9 greatly expands the genomic scope of *in vivo* editing, especially via viral vector delivery. The Nme2Cas9 all-in-one AAV delivery platform established in this study can in principle be used to target as wide a range of sites as SpyCas9 (due to the identical densities of optimal N_4CC and NGG PAMs), but without the need to deliver two separate vectors to the same target cells. The availability of a catalytically dead version of Nme2Cas9 (dNme2Cas9) also promises to expand the scope of applications such as CRISPRi, CRISPRa, base editing, and related approaches (Dominguez, Lim and Qi, 2016; Komor *et al.*, 2016) Moreover, Nme2Cas9's hyper-accuracy enables precise editing of target genes, potentially ameliorating safety issues resulting from off-target activities. Perhaps counterintuitively, the higher target site density of Nme2Cas9 (compared to that of Nme1Cas9) does not lead to a relative increase in off-target editing for the former. Similar results have been reported recently with SpyCas9

variants evolved to have shorter PAMs (Hu et al., 2018). Type II-C Cas9 orthologs are generally slower nucleases *in vitro* than SpyCas9 (Ma et al., 2015; Mir et al., 2018); interestingly, enzymological principles indicate that a reduced apparent kcat (within limits) can improve on- vs. off-target specificity for RNA-guided nucleases (Bisaria et al., 2017).

The discovery of Nme2Cas9 and Nme3Cas9 hinged on unexplored Cas9s that are highly related (outside of the PID) to an ortholog that was previously validated for human genome editing (Esvelt et al., 2013; Hou et al., 2013; Lee et al., 2016; Amrani et al., 2018). The relatedness of Nme2Cas9 and Nme3Cas9 to Nme1Cas9 brought an added benefit, namely that they use the exact same sgRNA scaffold, circumventing the need to identify and validate functional tracrRNA sequences for each. In the context of natural CRISPR immunity, the accelerated evolution of novel PAM specificities could reflect selective pressure to restore targeting of phages and MGEs that have escaped interference through PAM mutations (Deveau et al., 2008; Paez-Espino et al., 2015). Our observation that AcrIIC5$_{Smu}$ inhibits Nme1Cas9 but not Nme2Cas9 suggests a second, non-mutually-exclusive basis for accelerated PID variation, namely evasion of anti-CRISPR inhibition. We also speculate that accelerated variability may not be restricted to PIDs, perhaps resulting from selective pressures to evade anti-CRISPRs that bind other Cas9 domains. Cas9 inhibitors such as AcrIIC1 that bind more conserved regions of Cas9 likely present fewer routes toward mutational escape and therefore exhibit a broader inhibitory spectrum (Harrington et al., 2017a). Whatever the sources of selective pressure driving Acr and Cas9 co-evolution, the availability of

validated inhibitors of Nme2Cas9 (e.g. AcrIIC1-4) provides opportunities for additional levels of control over its activities.

The approach used in this studycan be implemented elsewhere, especially with bacterial species that are well-sampled at the level of genome sequence. This approach could also be applied to other CRISPR-Cas effector proteins such as Cas12 and Cas13 that have also been developed for genome or transcriptome engineering and other applications. This strategy could be especially compelling with Cas proteins that are closely related to orthologs with proven efficacy in heterologous contexts (e.g. in eukaryotic cells), as was the case for Nme1Cas9. The application of this approach to meningococcal Cas9 orthologs yielded a new genome editing platform, Nme2Cas9, with a unique combination of characteristics (compact size, dinucleotide PAM, hyper-accuracy, single-AAV deliverability, and Acr susceptibility) that promise to accelerate the development of genome editing tools for both general and therapeutic applications.

4.4 Materials and methods

Discovery of Cas9 orthologs with differentially diverged PIDs

Nme1Cas9 peptide sequence was used as a query in BLAST searches to find all Cas9 orthologs in *Neisseria meningitidis* species. Orthologs with >80% identity to Nme1Cas9 were selected for the remainder of this study. The PIDs were then aligned with that of Nme1Cas9 (residues 820-1082) using ClustalW2 and those with clusters of mutations in the PID were selected for further analysis. An unrooted phylogenetic tree of NmeCas9 orthologs was constructed using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Purification of Cas9 and Acr orthologs

Purification was performed as described in previous chapters. Briefly, Rosetta (DE3) cells containing the respective Cas9 plasmids were grown at 37°C to an OD_{600} of 0.6 and protein expression was induced by 1mM IPTG for 16 hr at 18°C. Cells were harvested and lysed by sonication in lysis buffer [50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 5 mM imidazole, 1 mM DTT] supplemented with 1 mg/mL Lysozyme and protease inhibitor cocktail (Sigma). The lysate was then run through a Ni^{2+}-NTA agarose column (Qiagen), and the bound protein was eluted with 300 mM imidazole and dialyzed into storage buffer [20 mM HEPES-NaOH (pH 7.5), 250 mM NaCl, 1 mM DTT]. For Acr proteins, 6xHis-tagged proteins were expressed in *E. coli* strain BL21 Rosetta (DE3). Cells were grown at 37 °C to an optical density (OD_{600}) of 0.6 in a shaking incubator. The bacterial cultures were cooled to 18°C, and protein expression was induced by adding 1 mM IPTG for overnight expression. The next day, cells were harvested and resuspended in lysis buffer supplemented with 1 mg/mL Lysozyme and protease inhibitor cocktail (Sigma) and protein was purified using the same protocol as for Cas9. The 6xHis tag was removed by incubation of the resin-bound protein with Tobacco Etch Virus (TEV) protease overnight at 4°C to isolate untagged Acrs.

Transfections and mammalian genome editing

Human codon-optimized Nme2Cas9 was cloned by Gibson Assembly into the pCDest2 plasmid backbone previously used for Nme1Cas9 and SpyCas9 expression (Amrani *et al.*, 2018a; Pawluk *et al.*, 2016). Transfection of HEK293T and HEK293T-TLR2.0 cells was performed as previously described (Amrani *et al.*, 2018a). For Hepa1-6 transfections, Lipofectamine LTX was used to transfect 500ng of all-in-one AAV.sgRNA.Nme2Cas9 plasmid in 24-well plates (~10^5 cells/well), using cells that had been cultured 24 hours before transfection. For K562 cells stably expressing Nme2Cas9 delivered via lentivector (see below), 50,000 – 150,000 cells were electroporated with 500 ng sgRNA plasmid using 10 μL Neon tips.

Lentiviral transduction of K562 cells to stably express Nme2Cas9

K562 cells stably expressing Nme2Cas9 were generated as previously described for Nme1Cas9 (Amrani *et al.*, 2018a). For lentivirus production, the lentiviral vector was co-transfected into HEK293T cells along with the packaging plasmids (Addgene 12260 & 12259) in 6-well plates using TransIT-LT1 transfection reagent (Mirus Bio). After 24 hours, culture media was aspirated from the transfected cells and replaced with 1 mL of fresh DMEM. The next day, the supernatant containing the virus was collected and filtered through a 0.45 μm filter. 10 uL of the undiluted supernatant along with 2.5 ug of Polybrene was used to transduce ~10^6 K562 cells in 6-well plates. The transduced cells were selected using media supplemented with 2.5 μg/mL puromycin.

RNP Delivery for mammalian genome editing

For RNP experiments, the Neon electroporation system was used exactly as described (Amrani *et al.*, 2018a). Briefly, 40 picomoles of 3xNLS-Nme2Cas9 along with 50 picomoles of T7-transcribed sgRNA was assembled in buffer R and electroporated using 10 μL Neon tips. After electroporation, cells were plated in pre-warmed 24-well plates containing the appropriate culture media without antibiotics. Electroporation

parameters (voltage, width, number of pulses) were 1150 V, 20 ms, 2 pulses for HEK293T cells; 1000 V, 50 ms, 1 pulse for K562 cells.

GUIDE-seq

GUIDE-seq experiments were performed as described previously (Tsai *et al.*, 2014), with minor modifications (Bolukbasi *et al.*, 2015). Briefly, HEK293T cells were transfected with 200 ng of Cas9 plasmid, 200 ng of sgRNA plasmid, and 7.5 pmol of annealed GUIDE-seq oligonucleotides using Polyfect (Qiagen). Alternatively, Hepa1-6 cells were transfected as described above. Genomic DNA was extracted with a DNeasy Blood and Tissue kit (Qiagen) 72 h after transfection according to the manufacturer's protocol. Library preparation and sequencing were performed exactly as described previously (Bolukbasi *et al.*, 2015). For analysis, all sequences with up to ten mismatches with the target site, as well as a C in the fifth PAM position (N_4CN), were considered potential off-target sites. Data were analyzed using the Bioconductor package GUIDEseq version 1.1.17 (Zhu *et al.*, 2017).

Targeted deep sequencing and analysis

We used targeted deep sequencing to confirm the results of GUIDE-seq and to measure indel rates with maximal accuracy. We used two-step PCR amplification to produce DNA fragments for each on- and off-target site. For SpyCas9 editing at DS2 and DS6, we selected the top off-target sites based on GUIDE-seq read counts. For SpyCas9 editing at DS4, fewer candidate off-target sites were identified by GUIDE-seq, and only those with NGG (DS4|OT1, DS4|OT3, DS4|OT6) or NGC (DS4|OT2) PAMs were examined by sequencing. In the first step, we used locus-specific primers bearing universal overhangs with ends complementary to the adapters. In the first step, 2x PCR master mix (NEB) was used to generate fragments bearing the overhangs. In the second step, the purified PCR products were amplified with a universal forward primer and indexed reverse primers. Full-size products (~250 bp) were gel-purified and sequenced on an Illumina MiSeq in paired-end mode. MiSeq data analysis was performed as previously described (Ibraheim *et al.*, 2018a; Pinello *et al.*, 2016).

Off-target analysis using CRISPRseek

Global off-target predictions for TS25 and TS47 were performed using the Bioconductor package CRISPRseek. Minor changes were made to accommodate characteristics of Nme2Cas9 not shared with SpyCas9. Specifically, we used the following changes to: gRNA.size = 24, PAM = "NNNNCC", PAM.size = 6, RNA.PAM.pattern = "NNNNCN", and candidate off-target sites with fewer than 6 mismatches were collected. The top potential off-target sites based on the numbers and positions of mismatches were selected. Genomic DNA from cells targeted by each respective sgRNA was used to amplify each candidate off-target locus and then analyzed by TIDE.

Mouse strains and embryo collection

All animal experiments were conducted under the guidance of the Institutional Animal Care and Use Committee (IACUC) of the University of Massachusetts Medical School. C57BL/6NJ (Stock No. 005304)
mice were obtained from The Jackson Laboratory. All animals were maintained in a 12 h light cycle. The middle of the light cycle of the day when a mating plug was observed was considered embryonic day 0.5 (E0.5) of gestation. Zygotes were collected at E0.5 by tearing the ampulla with forceps and incubation in M2 medium containing hyaluronidase to remove cumulus cells.

In vivo AAV8.Nme2Cas9+sgRNA delivery and liver tissue processing

For the AAV8 vector injections, 8-week-old female C57BL/6NJ mice were injected with 4 x10^{11} genome copies per mouse via tail vein, with the sgRNA targeting a validated site in either *Pcsk9* or *Rosa26*. Mice were sacrificed 28 days after vector administration and liver tissues were collected for analysis. Liver tissues were fixed in 4% formalin overnight, embedded in paraffin, sectioned and stained with hematoxylin and eosin (H&E). Blood was drawn from the facial vein at 0, 14 and 28 days post injection, and serum was isolated using a serum separator (BD, Cat. No. 365967) and

stored at -80°C until assay. Serum cholesterol level was measured using the Infinity™ colorimetric endpoint assay (Thermo-Scientific) following the manufacturer's protocol and as previously described (Ibraheim *et al.*, 2018a). For the anti-PCSK9 Western blot, 40 μg of protein from tissue or 2 ng of Recombinant Mouse PCSK9 Protein (R&D Systems, 9258-SE-020) were loaded onto a MiniPROTEAN® TGX™ Precast Gel (Bio-Rad). The separated bands were transferred onto a PVDF membrane and blocked with 5% Blocking-Grade Blocker solution (Bio-Rad) for 2 hours at room temperature. Next, the membrane was incubated with rabbit anti-GAPDH (Abcam ab9485, 1:2,000) or goat anti-PCSK9 (R&D Systems AF3985, 1:400) antibodies overnight. Membranes were washed in TBST and incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit (Bio-Rad 1706515, 1:4,000), and donkey anti-goat (R&D Systems HAF109, 1:2,000) secondary antibodies for 2 hours at room temperature. The membranes were washed again in TBST and visualized using Clarity™ western ECL substrate (Bio-Rad) using an M35A XOMAT Processor (Kodak).

Ex vivo AAV6.Nme2Cas9 delivery in mouse zygotes

Zygotes were incubated in 15 ul drops of KSOM (Potassium-Supplemented Simplex Optimized Medium, Millipore, Cat. No. MR-106-D) containing $3x10^9$ or $3x10^8$ GCs of AAV6.Nme2Cas9.sg*Tyr* vector for 5-6 h (4 zygotes in each drop). After incubation, zygotes were rinsed in M2 and transferred to fresh KSOM for overnight culture. The next day, the embryos that advanced to 2-cell stage were transferred into the oviduct of pseudopregnant recipients and allowed to develop to term.

CHAPTER 5: Therapeutic Genome Editing for Amyotrophic Lateral Sclerosis

5.1 Introduction

Amyotrophic lateral sclerosis (ALS) is a debilitating motor neuron disease affecting 3 in 100,000 adults in the United States. It is progressive, with only ~50% of patients surviving >30 months after symptoms emerge (Taylor, Brown and Cleveland, 2016; Zarei *et al.*, 2015). Although ALS can manifest in different ways, most patients experience motor neuron dysfunction, joint subluxation, and muscle loss (Zarei *et al.*, 2015; Kiernan *et al.*, 2011). There is currently no cure for ALS, and only two FDA-approved drugs are available for managing symptoms. Although these drugs have certainly helped suffering patients, their effect on prolonging survival is moderate (i.e. a few months) (Petrov et al., 2017). Alternative therapeutic strategies that target underlying causes of this disease may be more effective at preventing ALS progression.

Although the cause of most ALS cases is unknown (termed "sporadic"), ~10% of cases are inherited (termed "familial"). Moreover, mutations in ~20 genes have been associated with both ALS forms (Taylor, Brown and Cleveland, 2016). Among these genes, mutations in *C9ORF72* account for the majority of familial ALS cases and ~5% of sporadic ALS cases. Specifically, the intronic GGGGCC repeat expansion in *C9ORF72* is found in some ALS patients, with disease severity directly correlating with the number of repeats (Haeusler, Donnelly and Rothstein, 2016). Evidence has shown that this repeat expansion reduces levels of C9ORF72, suggesting that haploinsufficiency triggers ALS (Shi *et al.*, 2018). However, others have observed that repeat-expanded *C9ORF72* RNA transcripts form nuclear foci, perturbing normal

nuclear processes (Zu *et al.*, 2013; Peters *et al.*, 2015). Finally, cell culture data suggest that abnormal protein translation caused by the repeat expansion, known as repeat-associated non-ATG (RAN) translation, produces toxic dipeptides that form aggregates (Freibaum and Taylor, 2017). Although the exact mechanism underlying the C9ORF72 repeat expansion in ALS is unclear, all three processes likely contribute to disease progression.

Eliminating the intronic GGGGCC repeat expansion in *C9ORF72* at the DNA level may be a promising therapeutic approach for *C9ORF72*-associated ALS. AAV delivery of CRISPR-Cas9 with two sgRNAs targeting the flanking region is a viable approach. However, the restrictive PAMs of small Cas9 orthologs hinder cleavage close to the expanded repeats (Figure 5.1). An inability to cleave close to the repeats leads to excision of flanking exons, which may impair the activity or expression of C9ORF72. Excision of the *C9ORF72* repeat expansion using the hyper-accurate Nme2Cas9 may be a more clinically-relevant approach for ALS treatment. Firstly, its small size enables packaging into an all-in-one AAV vector, even with two gRNAs. Secondly, the location of Nme2Cas9 PAM sites in the target intron enables precise excision of repeats without breaching the flanking exons (Figure 5.1A).

5.2 Results:

5.2.1 Validating guides that efficiently excise the C9ORF72 repeats

Efficient targeting by two gRNAs flanking the repeat expansion is critical to the success of the deletion approach. To this end, we designed six sgRNAs targeting wt C9ORF72 in HEK293T cells. Although wild-type C9ORF72 does not possess expanded repeats, the flanking sequence is identical and functional sgRNAs should work on expanded C9ORF72. We cloned each candidate sgRNA into a U6-driven plasmid and transiently transfected them, along with a plasmid expressing Nme2Cas9, into HEK293T cells. We harvested cells after 72 hours, extracted genomic DNA, and measured indels using TIDE. While all sites were edited, the most efficient sgRNAs targeted TS2 and TS6 on either end of the repeats (Figure 5.1B). Next, we transfected both plasmids expressing each sgRNA along with Nme2Cas9 to test for the excision efficiency. Our results showed that the repeats were excised in 53% of cells, suggesting Nme2Cas9 can efficiently eliminate the repeats (Figure 5.1C). These two sgRNAs will be used for the remainder of this study.

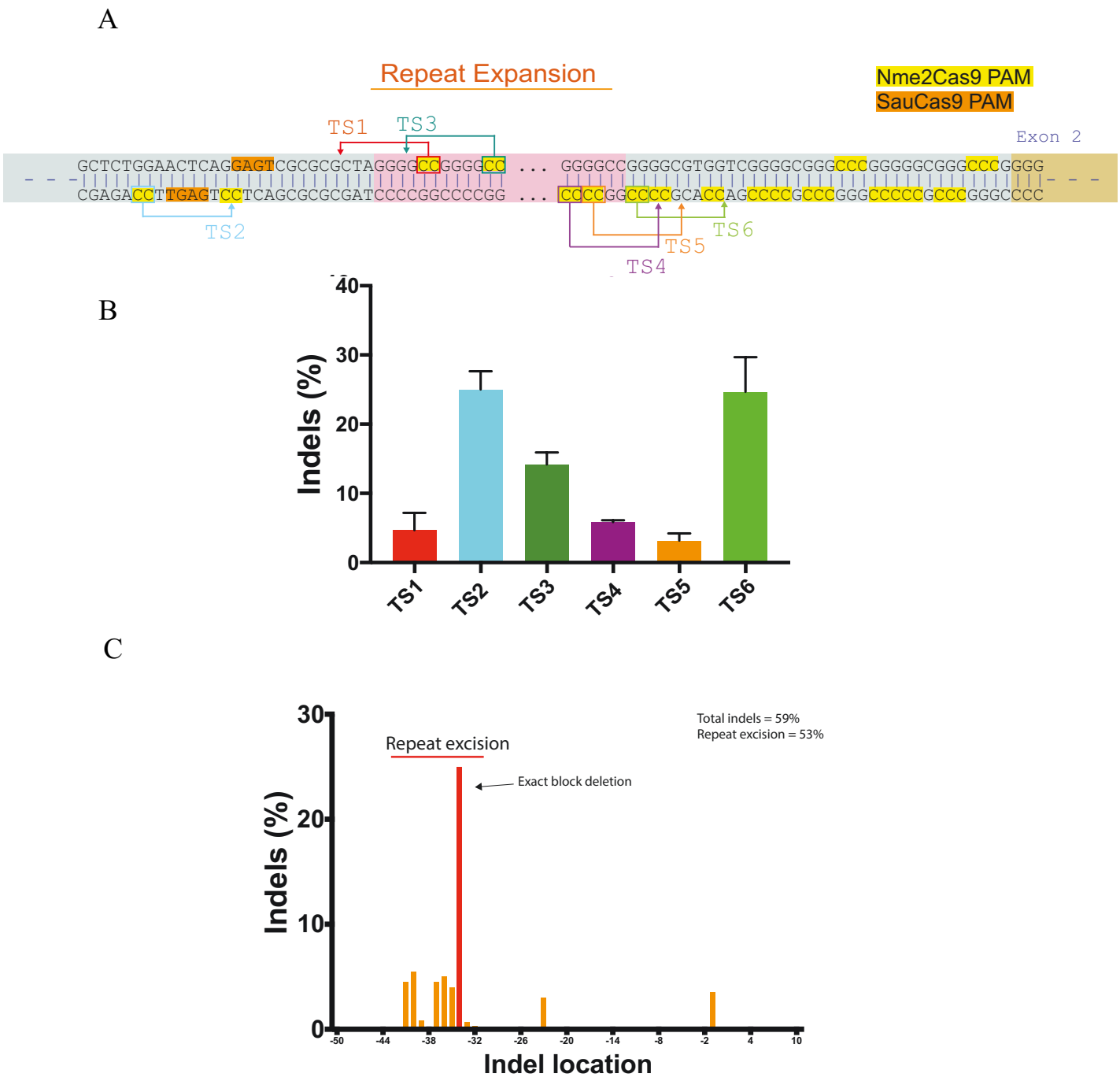


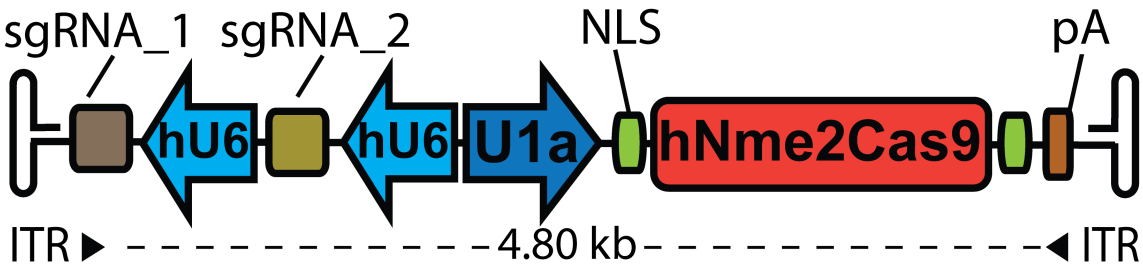
Figure 5.1 Efficient editing at sites flanking the repeats by Nme2Cas9.

(A) Six sites targeted by Nme2Cas9 flanking the repeat expansion. (B) Each site is targeted in HEK293T cells by transient transfection along with a plasmid expression Nme2Cas9. Indels measured by TIDE (%) are shown for each site. (C) Indel spectrum from TS2 and TS6 cotransfections in HEK293T cells measured by TIDE. The red bar signifies the expected excision of repeats.

5.2.2 Validation of dual-gRNA single AAV for excising C9ORF72 repeats

Cloning two sgRNAs along with Nme2Cas9 into a single AAV vector is imperative to the success of our deletion strategy. Our lab has now developed a more compact AAV backbone that allows for the expression of two sgRNAs along with their respective U6 promoters. This backbone is under 5000 nt in length which makes it ideal for packaging (Figure 5.2A). We cloned our top two guides into this vector and transiently transfected HEK293T cells. As a control, we transfected the 3 plasmids individually to compare excision levels. 72 hours after transfection, we isolated gDNA from each sample and analyzed the indel spectrum using TIDE. Oure data suggest that the dual-sgRNA AAV system is almost as efficient in excising repeats as the three plasmids separately being introduced (Figure 5.2B).

A



B

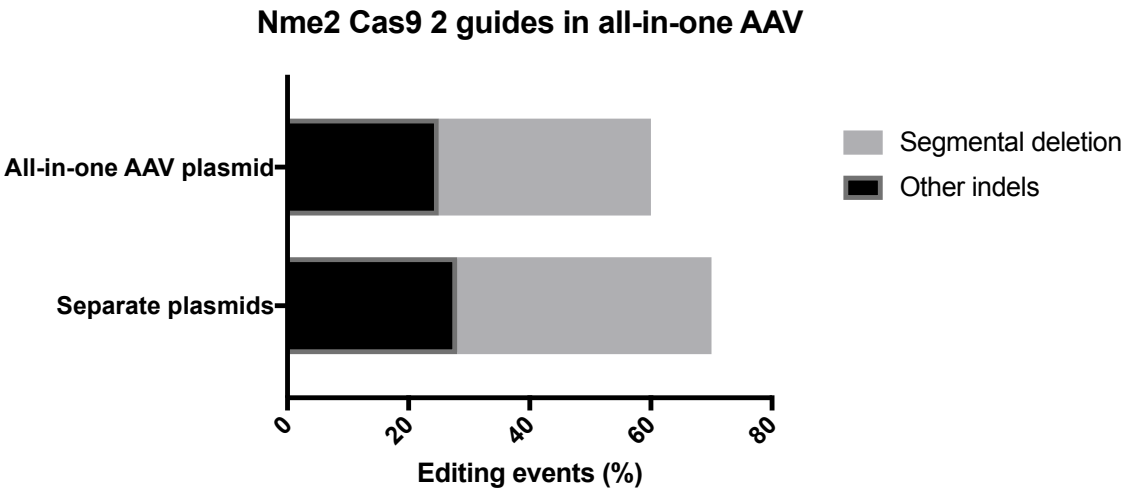


Figure 5.2 Validation of dual-sgRNA all-in-one AAV for excising C9ORF72 repeats

(A) All-in-one-AAV vector consisting of two sgRNAs and Nme2Cas9. Each sgRNA is driven by a U6 promoter, and Cas9 is driven by the U1a promoter.

(B) Comparison in segmental deletion i.e. excision of the repeats and other indels between the all-in-one AAV vector and separate plasmids.

5.2.3: A strategy for "collapsing" the repeats using a single guide RNA

While the dual guide strategy using Nme2Cas9 is highly promising, there are concerns regarding efficacy and the unpredictable nature of repair outcomes. While screening for efficient Nme2Cas9 sgRNAs targeting the flanking regions of the repeat, we noticed a peculiar indel pattern associated with TS3. Despite its low efficiency, the editing events all resulted in deletions in 6-nucleotide intervals (with ~ 6 or 12nt deletions). In other words, targeting within the repeats results in a "collapse" of the repeat on itself, resulting in incremental, 6-nucleotide deletions (not shown). This is in contrast to almost all other repair outcomes, where indels follow a random trajectory dominated by small insertions and deletions (Kosicki, Tomberg and Bradley, 2018). This could be in part due to the presence of repeats leading to repair by the microhomology-mediated end joining pathway (Iyer *et al.*, 2019). Microhomology-mediated end joining is an alternative form of the NHEJ pathway that uses microhomologous sequences during the alignment of ends being joined that results in precise deletions.

To find out if this phenomenon was unique to this site and/or to Nme2Cas9, we used SpyCas9 and targeted six sites within the repeats at either end of the GGGGCC repeats. These sites recognize NGG PAMs within the repeats, but the complementarity extends into the flanking region, thus providing a unique "handle" to provide specificity. We cloned each guide into a U6-driven plasmid and co-transfected HEK293T cells along with a plasmid expressing SpyCas9 under the CMV promotor. 72 hours post transfection, cells were isolated and indels were measured using TIDE.

While not every cleavage within the repeats resulted in a collapse, some led to efficient collapse at the expected 6-nt intervals. Of note, SpyTS3 resulted efficiently and consistently in collapse (Figure 5.3C).

It is important to note that while this finding warrants future investigation, it is possible that in the case of extended repeats, the collapse strategy may not work. For example, it may be the case that the three repeats in healthy mammalian cells results in the microhomology-mediated end joining pathway, but this may not be the case when more repeats are present. Nevertheless, proof-of-concept experiments on expanded repeats will answer these questions.

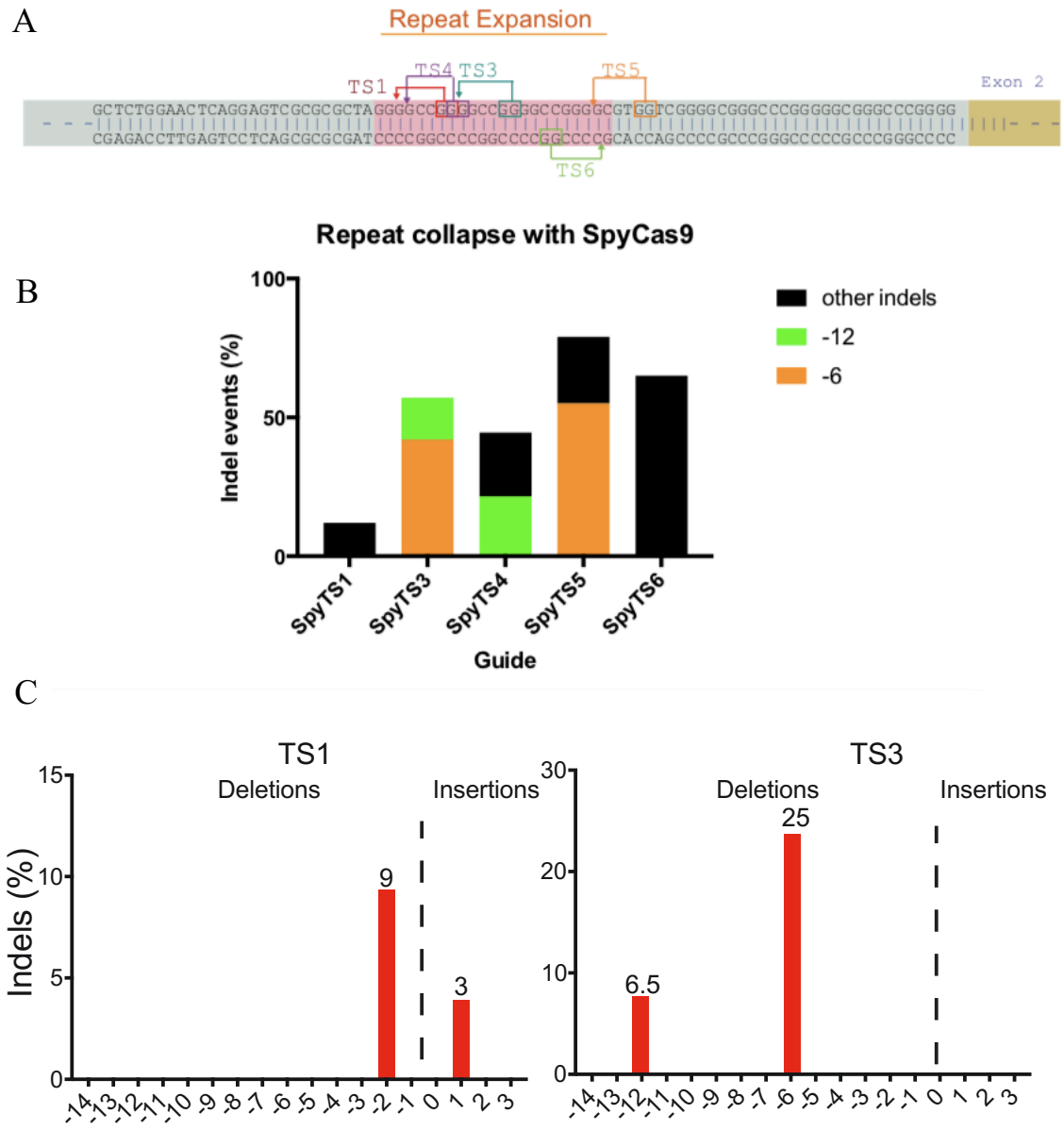


Figure 5.3. Testing the collapse strategy using SpyCas9
(A) Schematic of sites targeted by SpyCas9, flanking each end of the repeats.
(B) Repair outcome from five sites targeted with SpyCas9. Orange and green signify collapse while all other indels are in black.
(C) Indel spectrum of TS1 (shown as an example) compared to TS3, which is the site with repair outcomes entirely resulting in collapse.

5.2.4 Ongoing and future experiments

The preliminary studies presented here have paved the way to translate these findings to the context of ALS disease models. For the dual guide AAV approach, we will take advantage of cortical neurons from embryos expressing C9BAC [bacterial artificial chromosome containing parts of human C9ORF72 gene with ~500 repeats (Peters *et al.*, 2015)] and will infect them with AAV9.Nme2Cas9 along with both sgRNAs. The embryos will be E16 offspring from C9BAC homozygous parents and will possess the expanded repeats. Primary cortical neurons are easily cultured and are ideal for testing the feasibility of our approach in a neuronal cell type. For the collapse approach, the next step is to test the efficiency of collapse on expanded repeats.

The cellular phenotypes of C9ORF72-related ALS including RNA and dipeptide foci should be ameliorated if our approaches are successful. We will use RNA fluorescence in situ hybridization (RNA-FISH) to compare the formation of nuclear foci in treated vs control mice. Dipeptides produced from RAN translation form toxic cytoplasmic and nuclear aggregates in cells with expanded repeats. Among the six dipeptides formed in ALS pathology, sensitive rabbit monoclonal antibodies against poly glycine-proline (Poly-GP) dipeptides have been developed and used for detecting Poly-GP in patient samples. To quantify Poly-GP levels in treated vs. control groups, we will employ a sandwich immunoassay using the Meso Scale Discovery platform, an ELISA-based approach for detecting dipeptides (Ash *et al.*, 2013). If either approach is successful, we expect that Poly-GP quantities will be significantly lower in cells where C9ORF72 repeats were excised.

5.3 Discussion

ALS is a progressive motor neuron disease in which the average survival rate from onset is 2-4 years, and only 10% of individuals live past 10 years. There is currently no cure for ALS, and available medications do not significantly prolong survival. Mutations such as the repeat-expansion in C9ORF72 ultimately lead to neuronal death, decreased muscle mass, and loss of motor control. Gene editing approaches that eliminate ALS-causing mutations may prevent disease progression and improve the quality of life of some patients.

Our preliminary studies demonstrate CRISPR-Cas9's potential for the treatment of ALS caused by C9ORF72 repeat expansion. The next step will be to introduce the Nme2Cas9 dual sgRNA AAV9 vector into C9BAC neonates via facial vein injection. The goal is to offer hope for ALS patients and to get closer to the ultimate treatment for this disorder.

In addition to ALS, eliminating repeat expansions can also offer therapeutic advantages to other repeat-associated diseases. There are at least 40 different disorders associated with repeat expansions, most of which affect the central nervous system (Paulson, 2018). We envision that our single-AAV can be applied to many of these disorders as an *in vivo* therapeutic method.

5.4 Materials and methods

Plasmids used in this chapter

Nme2Cas9 plasmids were used from chapter 4. For the all-in-one AAV plasmid targeting the flanking region of C9ORF72, annealed oligos were sequentially cloned into plasmid pEJS1096. SpyCas9 lenti plasmid was obtained from Addgene (#86145). Annealed oligos were ligated after BsmbI digestion.

Transfections of HEK293T cells

Transfection of HEK293T and HEK293T-TLR2.0 cells was performed as previously described in chapter 4. For the experiments with separate plasmids, 200ng of Cas9 plasmid and 200ng of sgRNA plasmid were used. For SpyCas9 lentiviral constructs, 300 ng of the plasmid was used. To measure indels in all cells 72 hr after transfections, cells were harvested and genomic DNA was extracted using the DNaesy Blood and Tissue kit (Qiagen). The targeted locus was amplified by PCR, Sanger-sequenced (Genewiz), and analyzed by TIDE (Brinkman *et al.*, 2014) using the Desktop Genetics web-based interface (http://tide.deskgen.com).

Chapter 6: Conclusions and future directions.

6.1. The future of Cas9 orthologs and their inhibitors

There are currently thousands of sequenced Cas9 orthologs in the public database, of which only about a dozen have been carefully characterized, and a handful used for genome editing. With the current pace of sequencing platforms, and the recent explosion in microbiome and metagenomic sequencing, there will be thousands more sequenced within the coming decades. Each Cas9 ortholog comes with a unique set of features, and by characterizing more and more Cas9s, we are likely to uncover Cas9s with desired features. For example, KlaCas9 from the surface of a marine microalgae has a unique, NGGGGA PAM and is most efficient at lower temperatures, which could enable efficient genome editing in cases where experiments need to be performed at room temperatures or below. Such unique features will encourage researchers to characterize many more Cas9 orthologs.

The rate-limiting step in characterizing Cas9 orthologs is determining the PAM, and future studies should focus on streamlining the PAM identification process. The *in vitro* PAM identification strategy used in our studies is highly sensitive and identifies complex PAMs. Nevertheless, it requires that active protein is successfully purified and the sgRNA is correctly predicted (Briner *et al.*, 2014). The study of Cas9 diversity will be facilitated by the discovery of high-throughput PAM identification strategies. Alternatively, with the growth of artificial intelligence and our continued progress in understanding of protein:DNA interactions, one can envisage the possibility of predicting PAMs solely based on the PID peptide sequence. Nonetheless, as more and

more phage and other MGE sequences become available, we are more likely to be able to predict PAMs for a Cas9 of interest. For example, the PAM for HpaCas9 (N_4GATTT) could be predicted based on two hits in phages that infect *Haemophilus* species. While this is currently not possible with most Cas9s, the surge in sequencing will make this possible in the near future.

In addition to Cas9 diversity, Acr diversity remains to be fully appreciated. Future studies will take advantage of metagenomes to uncover new Acrs in a high-throughput fashion. In fact, a recent study took advantage of a systematic bioinformatics pipeline to reveal dozens of putative Acrs in diverse phyla (Watters *et al.*, 2018). Similar to Cas9, the promise of Acr applications must not be overstated, and we must be cautious before implementing Acrs.

6.2 The future of CRISPR genome engineering

The CRISPR genome editing toolbox has been expanding for the past decade. Clinical applications of CRISPR will continue to grow and the innovation of new technologies will continue to improve our understanding of biology and medicine. While Cas9 continues to be the most widely used CRISPR enzyme, other CRISPR nucleases have been introduced over the past 5 years that offer complementary features to Cas9. For example, the discovery of Cas12 and Cas13 nucleases has expanded the CRISPR toolbox (Shmakov *et al.*, 2017a). Cas12 is a DNA nuclease like Cas9, but it recognizes more AT rich PAMs, has the ability to process its own guide RNA (which enable multiplexing) and creates DNA cuts with overhangs (Zetsche *et al.*, 2015). Cas13 is another effector nuclease but instead of DNA, it is an RNA-guided RNA cutter (East-Seletsky *et al.*, 2016; Abudayyeh *et al.*, 2016; Abudayyeh *et al.*, 2017). Cas13 has been used for RNA editing and knockdown in human cells. It also exhibits collateral RNA shredding, i.e. it cleaves other RNAs in addition to the target RNA once activated, which has made it a powerful diagnostic tool (Gootenberg *et al.*, 2018). Undoubtedly, more CRISPR tools will be added in the near future, and more and more will near clinical and diagnostic applications.

However, it is important to not overstate the promise of CRISPR tools. History tells us that such breakthroughs, while powerful, often create a burst of hope among scientists and the public, but it takes decades to realize the complexity and nuances. For example, the discovery of RNA interference to silence RNAs appeared to be the magic bullet in the early 2000s. However, it took close to two decades for the first RNAi drug

to be approved by the FDA. CRISPR has certainly revolutionized biomedical sciences, but its future clinical implementation must be taken one step at a time. Nme2Cas9 is no exception, and below we describe future steps that will ensure that Nme2Cas9 can be safely and efficiently used.

6.2.1 The future of Nme2Cas9

Nme2Cas9 is a powerful genome editing tool that will streamline many genome editing applications. Our proof-of-concept *in vivo* delivery of a single-AAV Nme2Cas9 vector means that it can now be used for therapeutic purposes. The ability of Nme2Cas9 to precisely edit the gene of interest is of paramount importance in therapeutic applications, where the ideal number of off-targets is zero. However, there are several questions that can be addressed in future studies to improve Nme2Cas9's applications, some of which are outlined below

The safety of Nme2Cas9 must be investigated in detail. Considering that *Neisseria* species are human commensals/pathogens, the presence of antibodies against Nme2Cas9 is a possibility. Furthermore, while we have studied the specificity of Nme2Cas9 over a month-long period in mouse models, longer studies are needed to ensure safety in the long term. In addition to accuracy, cleavage efficiency is another feature of an ideal Cas9. Our comparison of Nme2Cas9 and SpyCas9 suggests that SpyCas9 is more efficient in our short-term, transfection-based experiments. While increased efficiency may come at an accuracy cost, it may be possible to engineer Nme2Cas9 variants with higher activity that retain their accuracy.

The development of fully modified crRNAs and tracrRNA (containing no 2′-OH groups) will significantly enhance Nme2Cas9 applications. Modified RNAs have been developed for RNAi and for SpyCas9 genome editing, and in both cases have increased the stability of the RNA and its efficacy *in vivo* (Mir *et al.*, 2018a; Khvorova and Watts, 2017; Hendel *et al.*, 2015). Chemical modifications such as 2′-O-methyl,

phosphorothioate and fluoro RNA have previously been used to synthesize guide RNAs for SpyCas9 with superior efficacy and even improved specificity. The development of such modified RNAs for Nme2Cas9 will enable *in vivo* delivery of RNP complex with the possibility of prolonged expression. The modified RNA also could ameliorate a potential immune response. The crystal structure of Nme2Cas9 can be used to introduce modifications in a structure-guided manner.

Cas9-based applications extend beyond its genome editing capabilities. While we demonstrated that Nme2Cas9 can be used as a nickase, there are applications that take advantage of Cas9's ability to guide other proteins to a given genomic locus. For example, our lab has fused dSpyCas9 to an engineered ascorbate peroxidase (APEX2) for spatially restricted enzymatic tagging of proteins associated with a locus of interest (Gao et al. 2018). One can envision the use of Nme2Cas9-Apex2 fusion to take advantage of the higher accuracy of Nme2Cas9, perhaps offering a higher signal to noise ratio. Other examples include fusing transcriptional activators/repressors, HDR effectors for improved homology repair, and base editors.

In this thesis, we present new Cas9 orthologs and Acrs to improve existing genome editing tools. The simplicity of Cas9s such Nme2Cas9 make it easier than ever to manipulate the human genome. However, our knowledge concerning the short- and long-term outcomes of Cas9 expression is extremely limited. Consequently, until preliminary animal studies have been completed, it is unequivocally dangerous to test these systems in humans. Taken together, although CRISPR is a powerful tool, we must exercise extreme caution in implementing it.

APPENDIX

Appendix 1: Information related to anti-CRISPRs and Cas9s in chapters 2 and 3

	Nucleotide sequence (5'→3')	Amino acid sequence (N→C)
AcrIE2	ATGGCCAATACCTATCTCATCGACCCCCGCAAAAACAACGACA ACTCCGGCGAGCGCTTCACGGTTGACGCTGTCGACATTACAGC CGCCGCGAAGAGCGCAGCCCAACAGATTCTTGGCGAGGAATTC GAGGGCCTCGTATACCGTGAAACCGGGGAGAGTAACGGAAGT GGCATGTTCCAGGCCTACCACCACCTGCACGGCACTAACCGCA CGGAGACGACCGTTGGCTATCCGTTTCATGTAATGGAACTCTG A	MANTYLIDPRKNNDNSGER FTVDAVDITAAAKSAAQQI LGEEFEGLVYRETG ESNGSGMFQAYHHLHGTN RTETTVGY PFHVMEL
AcrIIC1$_{Boe}$	ATGGCCAAGGAGGTCTTCAAGCTGAAGCCGGAGCTGGTGACGT ACAAGGGCTGCGGGTGGGCCCTGGCCTGCATCAAGGATGGCGA GATCATCGACCTGACCTACGTGCGTGACCTGGGCATCGAGGAG TACGATGAAAACTTCGACGGCCTGGAGCCGGAGATCATCTATT ACGACGTCGTCGCCTCGCAGGCGTGCAAGGAAGTGGCCTACCG CTATGAAGAGATGGGCGAATTCACCTTCGGCCTCTGCAGCTGC TGGGAATTCAACGTCATGTAA	MAKEVFKLKPELVTYKGCG WALACIKDGEIIDLTYVRDL GIEEYDENFDGLE PEIIYYDVVASQACKEVAY RYEEMGEFTFGLCSCWEFN VM
AcrIIC1$_{Nme}$	ATGGCCAATAAAACTTATAAAATTGGAAAAAATGCCGGGTATG ATGGCTGCGGTCTTTGTCTTGCGGCCATTTCTGAAAATGAAGC TATCAAAGTTAAGTATTTGCGCGACATTTGTCCTGATTACGAT GGCGATGATAAAGCTGAGGATTGGCTGAGATGGGGAACGGAC A GCCGCGTCAAAGCAGCCGCTCTTGAAATGGAGCAGTACGCATA TACGTCGGTTGGTATGGCCTCATGTTGGGAGTTTGTTGAACTA TGA	MANKTYKIGKNAGYDGCG LCLAAISENEAIKVKYLRDI CPDYDGDDKAEDWLRWGT DSRVKAAALEMEQYAYTS VGMASCWEFVEL
AcrIIC2$_{Nme}$	ATGGCCAGCAAAAACAATATTTTCAACAAGTATCCAACAATTA TTCACGGCGAAGCGCGGGGGGAGAATGACGAATTTGTGGTGCA TACGCGCTACCCGCGATTCTTGGCGCGGAAATCTTTTGACGAC AATTTCACGGGCGAAATGCCTGCAAAACCTGTTAACGGGGAAT TGGGACAAATCGGCGAACCGCGCCGCCTTGCTTATGATTCACG GCTTGGTTTGTGGCTTTCTGACTTCATCATGTTGGACAACAAC AAGCCGAAAAACATGGAGGATTGGCTTGGGCAATTAAAAGCC GCCTGCGATCGAATCGCGGCGGATGATTTGATGCTGAATGAAG ATGCGGCGGATTTGGAGGGCTGGGATGATTGA	MASKNNIFNKYPTIIHGEAR GENDEFVVHTRYPRFLARK SFDDNFTGEMPAKPVNGEL GQIGEPRRLAYDSRLGLWL SDFIMLDNNKPKNMEDWL GQLKAACDRIAADDLMLNE DAADLEGWDD
AcrIIC3$_{Nme}$	ATGGCCTTCAAACGCGCTATTATCTTCACTTCTTTCAACGGCT TTGAAAAAGTTTCTCGAACTGAAAAACGCCGCCTTGCCAAAAT CATCAATGCTCGAGTTTCCATCATCGACGAATACTTGAGAGCC AAAGACACCAACGCATCGCTTGACGGTCAGTACCGCGCTTTCT TGTTCAACGACGAATCGCCCGCAATGACCGAATTTCTGGCAAA ACTTAAAGCCTTTGCCGAAAGTTGCACCGGAATCAGCATCGAC GCATGGGAAATTGAAGAAAGCGAATACGTCCGCCTGCCGGTGG AACGCAGGGATTTCTTAGCGGCAGCCAACGGCAAAGAGATTTT TAAAATTTAA	MAFKRAIIFTSFNGFEKVSR TEKRRLAKIINARVSIIDEYL RAKDTNASLDGQYRAFLFN DESPAMTEFLAKLKAFAES CTGISIDAWEIEESEYVRLPV ERRDFLAAANGKEIFKI
AcrIIC4$_{Hpa}$	ATGAAGATCACCAGCAGCAACTTCGCGACCATTGCGACCAGCG AGAACTTTGCGAAGCTGAGCGTGCTGCCGAAAAACCACCGTGA GCCGATCAAGGGTCTGTTCAAAAGCGCGGTTGAACAGTTTAGC AGCGCGCGTGACTTCTTTAAGAACGAGAACTACAGCAAAGAGC TGGCGGAAAAGTTCAACAAAGAAGCGGTGAACGAGGCGGTTG AAAAGCTGCAAAAAGCGATCGATCTGGCGGAAAAACAGGGCA TTCAATTT	MKITSSNFATIATSENFAKL SVLPKNHREPIKGLFKSAVE QFSSARDFFKNENYSKELAE KFNKEAVNEAVEKLQKAID LAEKQGIQF
AcrIIC5$_{Smu}$	ATGAACAACAGCATCAAGTTCCACGTGAGCTACGACGGTACCG CGCGTGCGCTGTTTAACACCAAGGAGCAGGCGGAAAAATACTG CCTGGTTGAGGAAATTAACGATGAGATGAACGGCTATAAGCGT AAAAGCTGGGAGGAAAAGCTGCGTGAGGAAAACTGCGCGAGC GTGCAGGACTGGGTTGAGAAGAACTACACCAGCAGCTATAGCG ACCTGTTCAACATCTGCGAGATTGAAGTGAGCAGCGCGGGTCA ACTGGTTAAGATCGACAACACCGAGGTGGACGATTTCGTTGAA AACTGCTATGGCTTTACCCTGGAGGACGATCTGGAGGAATTCA ACAAGGCGAAACAGTACCTGCAAAAATTTTATGCGGAGTGCGA AAAC	MNNSIKFHVSYDGTARALF NTKEQAEKYCLVEEINDEM NGYKRKSWEEKLREENCAS VQDWVEKNYTSSYSDLFNI CEIEVSSAGQLVKIDNTEVD DFVENCYGFTLEDDLEEFN KAKQYLQKFYAECEN

HpaCas9	ATGGAAAATAAAAACTTAAATTATATCCTTGGTCTCGACCTTGG TATCGCCTCTGTTGGCTGGGCTGTGGTTGAAATTGACGAAAAA GAGAATCCTCTGCGTTTAATTGATGTTGGTGTACGTACTTTCGA ACGAGCAGAAGTGCCCAAAACAGGTGAAAGCCTTGCGCTTTCT CGCCGTTTAGCCCGCTCCGCTCGTCGATTAACTCAACGTCGCGT TGCTCGTCTGAAAAAAGCAAAACGACTTCTAAAATCAGAAAAT ATCCTATTATCCACTGATGAACGTCTTCCTCATCAAGTTTGGCA GCTACGTGTCGAAGGTCTAGATCATAAACTAGAACGCCAAGAA TGGGCTGCTGTGTTATTACATTTAATTAAGCACCGTGGTTATTT ATCTCAGCGTAAAAATGAAAGTAAAAGCGAAAATAAAGAGCT AGGTGCGTTATTAAGCGGTGTAGACAATAATCATAAATTACTG CAGCAAGCTACATATCGCTCGCCAGCAGAACTTGCCGTAAAAA AATTTGAGGTCGAAGAAGGTCATATCCGTAACCAACAAGGCGC CTATACTCACACGTTTAGCCGACTAGACTTACTCGCCGAAATGG AACTCCTCTTCTCTCGTCAACAACACTTTGGTAATCCATTCGCT TCAGAAAAATTATTAGAAAATTTGACCGCACTTTTAATGTGGC AAAAACCTGCCTTATCTGGTGAAGCTATTTTAAAAATGCTCGGT AAATGTACTTTTGAAGATGAATATAAGGCAGCTAAGAACACTT ATTCCGCAGAACGTTTTGTTTGGATAACGAAATTAAATAATTTA CGTATTCAAGAAAATGGCTTAGAGCGTGCTTTAAATGATAATG AACGTTTAGCATTAATGGAGCAACCTTATGACAAAAATAGATT ATTCTATTCACAAGTACGCTCAATATTAAAATTATCTGATGAGG CAATCTTTAAAGGCCTCCGTTATTCCGGTGAGGATAAAAAGGC CATTGAAACTAAAGCAGTACTGATGGAAATGAAAGCCTATCAC CAAATTCGTAAAGTATTGGAAGGTAATAACCTAAAAGCTGAAT GGGCAGAATTAAAGGCAAATCCAACATTATTAGATGAAATTGG TACGGCATTTTCATTATATAAAACTGATGAAGATATCAGCGCTT ACTTAGCAGGAAAACTCTCTCAGCCTGTATTAAATGCCTTATTG GAAAATCTTAGTTTTGATAAATTTATCCAATTATCCCTTAAAGC TTTATATAAACTTCTACCATTAATGCAACAAGGACTACGCTATG ATGAGGCTTGTCGTGAAATTTATGGCGATCATTATGGTAAAAA AACAGAAGAAAACCATCACTTCTTACCACAGATTCCAGCTGAT GAAATCCGTAATCCTGTGGTTTTACGAACACTCACGCAAGCTC GTAAAGTGATTAATGGCGTAGTGAGATTATATGGTTCGCCAGC TCGTATTCATATTGAAACGGGACGAGAAGTTGGCAAATCTTAC AAAGATCGTCGTGAACTAGAAAAACGTCAGGAAGAAAATCGT AAACAGCGTGAAAACGCAATCAAGGAATTTAAAGAGTACTTCC CTCATTTTGCTGGTGAGCCTAAAGCCAAAGACATTTTAAAAAT GCGACTTTATAAACAGCAAAATGCAAAATGTTTATATTCCGGC AAACCTATCGAATTGCACCGTTTATTAGAAAAAGGCTATGTAG AAGTCGATCACGCCTTGCCGTTTTCCCGTACTTGGGATGATAGT TTCAACAATAAAGTACTGGTGCTTGCTAATGAAAATCAAAACA AAGGCAATTTAACCCCTTTTGAATGGCTTGATGGTAAACATAA CAGCGAACGCTGGAGAGCGTTCAAAGCATTAGTTGAAACCAGT GCATTCCCTTATGCGAAAAAACAACGCATCCTAAGCCAAAAAC TTGATGAAAAGGGCTTTATTGAACGTAATTTAAATGATACGCG CTACGTTGCTCGTTTCTTATGTAATTTTATTGCAGATAATATGC ACTTAACAGGTGAAGGAAAACGAAAAGTATTTGCTTCCAACGG GCAAATTACCGCTTTACTTCGTAGCCGTTGGGGATTAGCAAAAT CACGGGAAGACAATGACCGCCATCACGCCTTAGATGCCGTTGT GGTTGCTTGCTCAACCGTCGCCATGCAGCAAAAAATCACACGT TTTGTTCGTTTTGAAGCTGGTGATGTATTCACTGGTGAACGAAT AGATCGTGAAACAGGTGAAATTATTCCATTACACTTCCCTACTC CATGGCAATTTTTCAAACAAGAAGTTGAAATCCGAATTTTTAGT GATAATCCTAAACTGGAACTAGAAAATAGATTGCCGGATCGTC CACAAGCCAATCATGAATTTGTCCAACCTCTGTTTGTATCTCGA ATGCCAACACGCAAAATGACCGGTCAAGGACACATGGAAACC GTGAAATCAGCGAAACGTCTCAATGAAGGGATAAGCGTGATTA AAATGCCACTCACTAAATTAAAATTAAAAGATTTAGAATTGAT GGTAAATCGTGAACGTGAAAAAGATCTTTATGATACTTTAAAA GCTCGTCTAGAGGCTTTTAATGATGATCCTGCTAAAGCGTTTGC TGAACCTTTTATAAAAAAAGGTGGGGCTATTGTTAAATCAGTA CGAGTAGAACAAATACAAAAATCTGGCGTATTAGTCCGTGAGG GGAATGGTGTCGCGGATAATGCCTCAATGGTGAGAGTGGATGT ATTCACTAAAGGTGGAAAATATTTCCTTGTGCCAATTTATACCT GGCAAGTCGCTAAGGGAATTTTACCTAATAAAGCAGCAACACA ATATAAAGATGAGGAAGATTGGGAAGTGATGGATAACTCTGCA ACTTTCAAATTTTCATTACACCCAAATGACTTAGTAAAATTAGT CACTAAAAAGAAAACCATTTTAGGTTACTTTAATGGACTTAATC GTGCCACAGGTAATATAGATATTAAAGAACATGATTTAGATAA	MENKNLNYILGLDLGIASV GWAVVEIDEKENPLRLIDV GVRTFERAEVPKTGESLALS RRLARSARRLTQRRVARLK KAKRLLKSENILLSTDERLP HQVWQLRVEGLDHKLERQ EWAAVLLHLIKHRGYLSQR KNESKSENKELGALLSGVD NNHKLLQQATYRSPAELAV KKFEVEEGHIRNQQGAYTH TFSRLDLLAEMELLFSRQQH FGNPFASEKLLENLTALLM WQKPALSGEAILKMLGKCT FEDEYKAAKNTYSAERFVW ITKLNNLRIQENGLERALND NERLALMEQPYDKNRLFYS QVRSILKLSDEAIFKGLRYS GEDKKAIETKAVLMEMKA YHQIRKVLEGNNLKAEWAE LKANPTLLDEIGTAFSLYKT DEDISAYLAGKLSQPVLNA LLENLSFDKFIQLSLKALYK LLPLMQQGLRYDEACREIY GDHYGKKTEENHHFLPQIP ADEIRNPVVLRTLTQARKVI NGVVRLYGSPARIHIETGRE VGKSYKDRRELEKRQEENR KQRENAIKEFKEYFPHFAGE PKAKDILKMRLYKQQNAK CLYSGKPIELHRLLEKGYVE VDHALPFSRTWDDSFNNKV LVLANENQNKGNLTPFEWL DGKHNSERWRAFKALVETS AFPYAKKQRILSQKLDEKG FIERNLNDTRYVARFLCNFI ADNMHLTGEGKRKVFASN GQITALLRSRWGLAKSRED NDRHHALDAVVVACSTVA MQQKITRFVRFEAGDVFTG ERIDRETGEIIPLHFPTPWQF FKQEVEIRIFSDNPKLELENR LPDRPQANHEFVQPLFVSR MPTRKMTGQGHMETVKSA KRLNEGISVIKMPLTKLKLK DLELMVNREREKDLYDTLK ARLEAFNDDPAKAFAEPFIK KGGAIVKSVRVEQIQKSGV LVREGNGVADNASMVRVD VFTKGGKYFLVPIYTWQVA KGILPNKAATQYKDEEDWE VMDNSATFKFSLHPNDLVK LVTKKKTILGYFNGLNRAT GNIDIKEHDLDKSKGKQGIF EGVGIKLALSFEKYQVDEL GKNIRLCKPSKRQPVR

	ATCAAAAGGGAAACAAGGTATTTTTGAAGGTGTTGGTATTAAA TTAGCCCTTTCCTTCGAAAAATATCAAGTCGATGAACTTGGAAA AAATATTCGTTTATGTAAACCAAGTAAACGCCAACCTGTTCGTT AA	
SmuCas9	ATGATGATGGAAAAATTTCACTATGTATTGGGTTTGGATTTGGG TATCGCCTCTGTGGGGTGGGCTGCCATTGAAATTGACAAGGAA ACCGAAACATCAATCGGTTTATTGGATTGCGGTGTCAGAACAT TTGAACGTGCAGAAGTACCCAAAACAGGCGATTCTCTTGCCAA AGCTCGCCGTGAAGCCAGAAGTACTCGCCGTTTAATTCGCAGA CGTTCGCATCGCTTATTACGTTTAAAACGTTTATTGAAACGTGA AATTTTCAGGCAGCCTGAAACGTTTAAAGACTTACCAATCAAT GCTTGGCAATTGCGTGTTAAAGGCTTGGATAGTCGGTTGAATG AATATGAATGGGCGGCCGTTTTATTGCATTTGGTGAAGCATCGC GGTTATTTATCGCAACGCAAAAGCGAAATGAGCGAAACAGACA GCAAATCTGAAATGGGCAGATTACTGGCAGGTGTGGCGGAAAA TCACCAACTTTTACAACAAGAACAATATCGTACACCAGCCGAA TTAGCACTCAAAAAATTTGTGAAACATTTTCGCAATAAAGGTG GCGATTATGCACACACTTTCAACCGTTTGGATTTGCAAGCCGAA TTGCATTTATTGTTTCAAAAACAACGTGAATTAGGCAATCCATT CACTTCACCAGAATTGGAACGGCAAGTTGATGATTTGTTGATG ACGCAGCGCAGTGCTTTACAAGGTGATGCGATTTTGAAAATGT TGGGTCATTGTGGGTTTGAACCTGAACAATTCAAAGCAGCGAA AAACACATTCAGTGCCGAACGTTTTATTTGGTTGACAAAACTCA ATAATCTTCGCATTCAAGACCAAGGCAAAGAACGTGCGTTAAC TGCCGATGAGCGTACCAAATTGTTGGACGAGCCTTATAAAAAA AGTAAATTGACTTACGCACAAGTTCGCAAATTATTAAGCTTGCC TCAAACTGCTATTTTTAAAGGTTTGCGTTATGATTTGGAACATG ACAAAAAAGCAGAAAACAGTACGTTGATGGAAATGAAATCCT ATCACAACATCCGCCAAACATTGGAAAAATCAGGTTTGAAAAC AGAATGGCAAAGTATTGCCACGCAGCCTGAAATTTTAGATGCA ATTGGCACGGCGTTTTCCATTTATAAAACCGATGAAGATATTTC GCATGAATTAAAAACGTGCAGGCTGCCTGAAAACGTATTGAAT GAATTACTGAAAAACATCAATTTTGATGGATTCATTCAATTATC GTTGACTGCATTACGCAAAATTTTGCCCTTGATGGAACAAGGCT ACCGTTATGATGAAGCGTGTACCCAAATTTACGGTAATCATCAT TCAGGCAGCTTGCAACAAGAATCAAAGCAATTTTTGCCACATA TTCCGATTGATGATGTCCGAAATCCTGTGGTGTTCCGTACTTTG ACCCAAGCAAGAAAAGTGGTGAATGCGATTATTCGTCGGTATG GTTCGCCAGCTCGTGTGCATATTGAGATGGCGCGTGAATTGGG TAAGTCTAAATCAGACCGTGACCGAATTGAAAAACAACAACAA AAAAATAAAAAAGAACGTGAAAACGCAGTCGCCAAATTCAAA GAAGATTTCCCTGATTTTGTGGGCGAACCCAGAGGGAAAGATA TTTTGAAAATGCGTTTGTATGAACAACAACACGGCAAATGTTT GTATTCGGGTCATGATATTGATATTAATCGATTGAATGAAAAA GGTTATGTTGAAATTGACCATGCCCTGCCATTTTCACGGACTTG GGATGATAGTCAAAATAATAAAGTTTTGGTACTTGGCAGCGAA AACCAAAATAAACGCAATCAAACGCCTGATGAATATTTGGACG GTGCAAACAATAGCCAACGTTGGCTTGAATTTCAAGCGCGTGT ACAAACTTGTCATTTTTCTTACGGTAAAAAACAACGCATTCAAT TAGCCAAATTAGACGATGAAACCGAAAAAGGATTTTTAGAACG CAATCTAAATGATACGCGGTATATTGCTCGTTTTATGTGTCAAT TTGTCCAAGAAAATTTATATTTGACAGGTAAAGGAAAACGTCT TGTTTTTGCATCAAACGGCGGAATGACCGCAACATTGAGAAAT TTATGGGGTTTGAGAAAAGTCCGTGAAGACAACGACCGCCATC ATGCTCTTGATGCGATTGTGGTGGCGTGTTCCACTGCTTCTATG CAACAAAAAATAACCAAAGCATTTCAACGCCATGAAAGCATTG AATATGTGGATACCGAAACGGGCGAAGTAAAATTTCGTATTCC ACAGCCTTGGGATTTTTTCCGTCAGGAAGTGATGATTCGTGTGT TCAGCGACCAACCGTGTGAAGATTTGGTAGAAAAATTGTCGGC TCGTCCCGAAGCTTTGCATGACAACGTAACGCCTTTATTTGTCT CGCGTGCACCAAATCGCAAAATGTCGGGGCAAGGGCATTTGGA AACCATCAAATCTGCAAAAAGGCTGTCTGAAGAAAACAGTATG GTTAAAAAACCATTAACCACATTGAAATTAAAAGATATTCCAG	MMMEKFHYVLGLDLGIAS VGWAAIEIDKETETSIGLLD CGVRTFERAEVPKTGDSLA KARREARSTRRLIRRRSHRL LRLKRLLKREIFRQPETFKD LPINAWQLRVKGLDSRLNE YEWAAVLLHLVKHRGYLS QRKSEMSETDSKSEMGRLL AGVAENHQLLQQEQYRTPA ELALKKFVKHFRNKGGDY AHTFNRLDLQAELHLLFQK QRELGNPFTSPELERQVDDL LMTQRSALQGDAILKMLGH CGFEPEQFKAAKNTFSAERF IWLTKLNNLRIQDQGKERA LTADERTKLLDEPYKKSKL TYAQVRKLLSLPQTAIFKGL RYDLEHDKKAENSTLMEM KSYHNIRQTLEKSGLKTEW QSIATQPEILDAIGTAFSIYK TDEDISHELKTCRLPENVLN ELLKNINFDGFIQLSLTALR KILPLMEQGYRYDEACTQI YGNHHSGSLQQESKQFLPHI PIDDVRNPVVFRTLTQARK VVNAIIRRYGSPARVHIEMA RELGKSKSDRDRIEKQQQK NKKERENAVAKFKEDFPDF VGEPRGKDILKMRLYEQQH GKCLYSGHDIDINRLNEKG YVEIDHALPFSRTWDDSQN NKVLVLGSENQNKRNQTPD EYLDGANNSQRWLEFQAR VQTCHFSYGKKQRIQLAKL DDETEKGFLERNLNDTRYI ARFMCQFVQENLYLTGKG KRLVFASNGGMTATLRNL WGLRKVREDNDRHHALDA IVVACSTASMQQKITKAFQ RHESIEYVDTETGEVKFRIP QPWDFFRQEVMIRVFSDQP CEDLVEKLSARPEALHDNV TPLFVSRAPNRKMSGQGHL ETIKSAKRLSEENSMVKKPL TTLKLKDIPEIVGYPSREPQL YAALKTRLETHDDDPIKAF AKPFYKPNKNGELGALVRS VRVKGVQNTGVMVHDGK GIADNATMVRVDVYTKAG KNYLVPVYVWQVAQGILP NRAVTSGKSEADWDLIDES FEFKFSLSRGDLVEMISNKG RIFGYYNGLDRANGSIGIRE HDLEKSKGKDGVHRVGVK TATAFNKYHVDPLGKEIHR CSSEPRPTLKIKSKK

	AAATCGTAGGCTACCCGAGTCGTGAACCTCAATTGTATGCCGC ATTGAAAACACGTTTAGAAACGCATGATGATGACCCAATTAAA GCCTTTGCCAAACCCTTTTACAAACCCAATAAAAATGGTGAATT GGGGGCGTTGGTTCGATCGGTGCGTGTGAAAGGTGTACAAAAT ACGGGTGTAATGGTTCATGATGGCAAAGGCATTGCCGATAATG CCACAATGGTTCGTGTTGATGTCTATACCAAAGCGGGCAAAAA TTACCTTGTTCCTGTGTATGTTTGGCAGGTGGCTCAAGGAATTT TGCCAAATCGGGCGGTTACTTCTGGCAAAAGTGAAGCAGATTG GGATTTAATTGATGAAAGTTTTGAATTTAAATTTTCGCTGTCTC GTGGGGATTTAGTGGAAATGATTAGCAATAAAGGAAGAATTTT TGGTTATTACAATGGGTTAGATCGTGCAAATGGAAGTATTGGG ATTCGTGAACATGATTTGGAAAAGTCCAAAGGAAAAGATGGTG TTCATCGTGTTGGCGTGAAAACCGCCACCGCATTCAACAAATA CCACGTTGACCCACTTGGTAAAGAAATTCATCGGTGTTCATCTG AACCACGCCCCACATTAAAAATCAAATCCAAGAAATAA	
BoeCas9	ATGAACAATAATGACATCCCCCAAGCTACAGATACTTTTTCAC AAACGCCTGATTTTGTCCCTCATGATCCTAAGATTCTAAAAAAA GAGATTGATTATGTGATTGGTTTAGACATTGGCATAGCGAGTAT TGGGTGGGCGGTGCTAGCACTAGATGAGGAAGATAAACCGATA GCGCTTATTGATCGTGGAGTGAGATTATTTGATAAGGCGGAAA CTAAAAAAGGTGACTCTCTCTCTTTGGCCAGACGTACGGCTAG GGGGCAGCGGCGTCGGATTGGTAGACGTGTCAAGCGACTTGAC TTAGCTCGGCAGACTTTAATTGAGCACGGGGTCTTGCAACCAA ATGACCTTAATAATAATCCGGATGGTAATGATTTTGTGGCCAAT AGGCCTGATAATGCTCCTGGTAATACGGTATGGGACTTACGTG TCAAAGGTCTAGACCATAAGCTAAAATCTCATGAATGGGCGGC GGTACTTTTACATTTAATTAAGCGAAGAGGGTATCTGTCACAA GGCAAAAGCGAGGCGAAAGCAGATGAGGAGACTGGTAAACTT TTAGCAGGTACAGCAGAAAACCATAAAAAACTAGCTAATGGTA ACTACAGAACACCCGCAGAGTTAGCTGTCAACACCTTTGCTAA GGAGTCTGGTCGAATAAGAAACCAAGCAGGTAGCTATCAGCAT ACATTTCTTAGGGAAGACCTACAAAAAGAACTTCAGTCACTCT TTGAAAACCAACGTAATTCTAAAAACCCATACGCAAATGAAAC ACTTGAAAAGGCCTTTGATGACTACTTAACGTGGCAAAAACCT GCATTGTCAGATGATGCATTACTTAAAATGGTGGGGCCTTGTAC CTTCGAAGACGGAGAAAAACGTGCAGCCAAAGCCACCTATACG GCAGAAAGATTTACCTGGCTTACTAAGCTGGCTAATCTTAAAA TCATCACAAACGGTGTATCTCGAGAGTTAACACCTGAAGAGCG AGACAAACTAAAAGAGTTACCTTTTGAGTTGACAAGCCTTACA TTTTCTCAAGTACGTAAAACCTTAGGTTTGGATGAAAATAGTCG CTTTAATTTAGCGAGTTACAAAAGCAAGAAAAAAAGCAAGAA AAACAAGCAAGATACACAAAAAGAAAAAACGAAGACTGTAAT TAAAGACGCTGAAAAAGAAACCTTCATGGAGTGTAAGTCCTAT CATGCTATACGTAAGGCGCTAGGAAAGACATTAGATTCAGAGT GGGAGGCAATCAAAAACAATCCAGACTTACTTGATTTGATTGG TACGACGTTTACGATTTATAAAACTGATAAAGCTATCCGTAAA GAACTAAAAAGTCATCTATCTCCTGAAGCCATTAATGCCCTTCA GAAAAATTTAAGTTTTACTAAGTTTTTACATGTTTCTCTCAAAA TGATTAAGAAAATCATGCCGTATCTTGAAGAAGGCATGAGATA TGATGAGGCCTGCAAAGAGGTTTATGGTAGTCATTATGGCAAT AGCACTCAAAGGCTTCTAAAACAATTTTTGCCGCCTATGGATA GCGAAGAAAATGTATTAAATGATGTAATTCGCAACCCGGTGGT GGCACGTGCCGTTAGTCAAACGCGCAAAGTAATCAATGCCATT ATTAGACGCTATGGTAGTCCGACCAAAGTTATTATCGAAACGG GACGAGATTTAGGTAAATCATTTCAAGAAAGAAAAGAAATTAA AGATCATCAAAAGGAAAATCGCAATCAAAAACTAGCTGCTCAA GAGGAATTCGAAAAACGATTTGGTGTATCACCAAATGGTGGTG ATCTGCTTAAAATGCGTCTTTACGAGCAGCAACAAGGCAAATG CTTGTATAGCGGAACGACTATAGAGTTGAATCAACTCTTAGAC AAAGGCTATGTTGAGATTGACCATGCACTGCCATTTAGTCGTAC GTGGGATGACAGTTTTAACAACAAAGTACTTGTAATGGGATCT GAAAATCAGAATAAAGGTAATAGAACCCCTTATGAATACTTAG ACGGAGAGCATGATAGTGCGCGCTGGCAAACATATGTAGCGTT AGTTAATGGTTCAAATATGTCTTATTTAAAGAAACAGCGCTTGT TGACTAAAGAGATTAAAGAAGATGAATGGAAAGCGCGAAATC TTAATGATACTCGCTATATTGCGATTTATTTGCGTGACTTTATTA AGGCAAATTTACTATTAAAAAGCAGCAAAGAGAATAACGTGCT CACTACGAATGGTCGAGTGACTTCACTATTGAGAGGCCGCTGG GGATTAGCTAAAGATCGTGAAGCTAACGCTCGTCATCATGCTA TGGATGCCGTAGTCGTGGCGTGTACTACGCCTTCATTTATTCAG	MNNNDIPQATDTFSQTPDF VPHDPKILKKEIDYVIGLDIG IASIGWAVLALDEEDKPIALI DRGVRLFDKAETKKGDSLS LARRTARGQRRRIGRRVKR LDLARQTLIEHGVLQPNDL NNNPDGNDFVANRPDNAP GNTVWDLRVKGLDHKLKS HEWAAVLLHLIKRRGYLSQ GKSEAKADEETGKLLAGTA ENHKKLANGNYRTPAELAV NTFAKESGRIRNQAGSYQH TFLREDLQKELQSLFENQRN SKNPYANETLEKAFDDYLT WQKPALSDDALLKMVGPC TFEDGEKRAAKATYTAERF TWLTKLANLKIITNGVSREL TPEERDKLKELPFELTSLTFS QVRKTLGLDENSRFNLASY KSKKKSKKNKQDTQKEKT KTVIKDAEKETFMECKSYH AIRKALGKTLDSEWEAIKN NPDLLDLIGTTFTIYKTDKAI RKELKSHLSPEAINALQKNL SFTKFLHVSLKMIKKIMPYL EEGMRYDEACKEVYGSHY GNSTQRLLKQFLPPMDSEE NVLNDVIRNPVVARAVSQT RKVINAIIRRYGSPTKVIIET GRDLGKSFQERKEIKDHQK ENRNQKLAAQEEFEKRFGV SPNGGDLLKMRLYEQQQG KCLYSGTTIELNQLLDKGY VEIDHALPFSRTWDDSFNN KVLVMGSENQNKGNRTPY EYLDGEHDSARWQTYVAL VNGSNMSYLKKQRLLTKEI KEDEWKARNLNDTRYIAIY LRDFIKANLLLKSSKENNVL TTNGRVTSLLRGRWGLAKD REANARHHAMDAVVVACT TPSFIQRITNYMKRKEANER LETTDPVTGEVKRVHFPKP WDGFRSEVLASIFRENPQEE IPAYRKLVLDQYLDVDSIKP LFVSRMPRRVVTGEGHEAT IRSAKRLKEGLSTVKKSIND LSKSDLERFVNKDREPKFY QILKDRLETIENLVNALNES KNTDEEAQIKNKLNEVKEA PVHKLNNHGMPTGPIVKSIR LLVTPKSGVGIRKGVANNG KMVRVDIFKKGESFYGVPI YAWQVAKGILPNKGVNSK KSENDWELMDESAEFQFSC

	CGCATCACCAACTATATGAAACGTAAAGAGGCTAACGAGCGTC TTGAGACCACCGATCCAGTCACGGGTGAAGTTAAACGAGTGCA TTTTCCTAAACCATGGGACGGTTTTAGATCAGAGGTATTAGCGA GTATTTTTAGAGAAAATCCTCAGGAGGAGATCCCCGCCTACAG AAAATTAGTTTTAGATCAATATTTGGATGTAGATTCTATTAAAC CATTATTCGTCTCGCGAATGCCTAGAAGAGTGGTAACCGGTGA GGGGCATGAAGCAACTATTCGATCTGCTAAAAGACTCAAAGAG GGACTCAGCACAGTTAAAAAATCAATTAATGATTTGTCTAAGT CTGATCTTGAGAGGTTCGTTAATAAAGATCGGGAACCAAAATT TTACCAAATTTTGAAGGATCGTCTTGAAACTATTGAAAATTTAG TTAATGCTCTAAATGAGTCTAAAAATACGGATGAAGAAGCTCA AATAAAAAATAAACTTAATGAAGTTAAAGAAGCACCCGTGCAT AAATTGAATAATCATGGTATGCCTACAGGTCCAATAGTAAAGT CCATTCGGTTGTTAGTTACTCCAAAATCAGGAGTAGGTATTAGA AAAGGGGTAGCTAATAATGGTAAGATGGTCAGGGTGGACATCT TTAAAAAAGGTGAGTCCTTTTATGGTGTTCCTATTTATGCGTGG CAGGTAGCTAAGGGGATATTGCCTAATAAAGGTGTCAATAGTA AAAAGTCGGAAAATGATTGGGAGCTTATGGATGAATCAGCTGA GTTTCAATTTTCTTGTTTTAAAAATGATTTGATTAAAATTGTGCC CAAAACTGGTGATCACATCTTCGGTTATTACGTAGGCTTTGATA GAGATAGACGAAGTTTTACACTGGCAAAACATGATAAGGGTTC TACGAATGGAGAGGGAGATAAATCTGGTACGTTTCGATTCAGT ATCACAACAATGGCAAAGGAAATTATCAAGTGCGAGATTGACG TCCTCGGCCATACAATAACTGAAGTAAAACATGAAAAAAGGTT GGATTTTAGTAAATCGAGCAAGAAAAAAGATTGA	FKNDLIKIVPKTGDHIFGYY VGFDRDRRSFTLAKHDKGS TNGEGDKSGTFRFSITTMAK EIIKCEIDVLGHTITEVKHEK RLDFSKSSKKKD
KlaCas9	ATGGGTGGTTTTTTTCCATACAGGTTGGCACTAGATGGTGGCAT AGCATCTTTTGGCTGGGCCATTTATGCGCTTGACCGCAACGGAG AACCTTGTGATCTCATTGACGGCAATGTGCTTCTTTTCCCGACG GCGATTGGTGCTTCGAACAGAAGGGAAAAACGCAGTGCACGA AAAACCAATGATCGCAAGCAGCTCAGGCTGAAAAAAATATGTG AATTCTTTGCCTCAGATTTTCAGGGTTATAATCGTGAAAAGGCC CCCGCAGATGTCCTGAATACCTCACCTTTCAGGCTGCGGGCGA ATGCTCTGGATCGGCAAGTCAGTTTAGCCGAATTGCAACGTAT ATGCCTTCATATGGCAAAACATCGGGGCTCTTCTGCCATCCGGA TGTCTGAAGATGAGGAGATTGCTAAAGAGGGAGCGGCAATTTC ATCCGCCTTTGCCCGGACACAGGTACAGATGCTGGAACGATCT GCTCGTACATACGGGGAATTGCTTTGGCGGATAGAAAAGGCAA ACAATCGAAAGTCTGGACAAGACACAACAAAAATCCGGATCA GCAAACAAGGTAACAAAGCGGGTGATGCCTATTCTCATTATCC ACTCCGGTTTTTAATCGAACAGGAATTTGACAAGATTATGCTTG CTCAGTCAGCTTTTCACGACGAACTGACTCAGGAAAAAATCAA TGCTCTCAAAGAAATTCTGTTTTTCGAACGTGAAACAATAACAC CGAAACCGGGGTACTGTTTATATATCGAAAACGAACGTCGGAT ACCAAAACTCAGCCGCCTGTTTCAACTGAAACGTATTTATGAG GAGGTCAATAACCTTCGGATCGATACTGCTGAGGGAATCCGTA TTCTAACCTTTATTGAACGGGATCTTCTGATCGATCAGTTAATG GCAGGTGTCACGCTTGGCAAGAAAAACACCAAGGAGTTTTTGA AACTGAAGGGTAAAGATAAACTCTATCTTGGAAAAAAAGACAC TCTTGTCATTAAACCCTATCCCTTTGACTGTCTCCTGGGACAGG ACGCCGTTCTGGGGGACAGATGGGCAGAGATGTCAGAGCAGC AGCAGGATGAGTTGCTCGATATTCTGGCGACTGTTTGGGATGA TACACGCGCGCATATCGCGTTGAAAAAACTGTTGAACTGCAAT GATGTGCTGGCGCGCGCAATTCTTGCGGTGCCACTTCCTTCGGG GTATGGCGCGATTGGCGCTTCTGCGACCCGCGAGATTATCGAT GTGCTGGAAGCAAATATTTGCTCTGCTACAGAAGCAGCCATGT TGGCGGGTTTAAAGGAACGGGCAATATTCGAGGGGGATTTGCT TCCCCGTCTGCCGTATTACGGTGAAATTCTGGGTAATCATGTTA GTTTTAGCGACAATGCTGGTAGCAGCCTTCCTTATCAGGAAAA AGTGTTCGGGCGCATTCCAAATATGGTGGTTCACCGTGCTCTCA TCCAGCTTCGGGGATTGGTAAATAGCATCATCAAAAAATATGG GCCACCCACTCAGATTCATCTGGAACTGGGGCGGGAACTGTCC ATGTCTGAACAGGAACGCGATGATATTCAGAAAACCAACGAAA ATAATCGCAAGGTTAATGATGTTATTCGCAACAAGCTTGTGAA AAATGAGGAAAAACCCAATCGATCCAACCTTATCCGCTATAAA CTGTGGGAACGGCAAAAAGGCACCTGTATCTACAGCGGGGACA ATATAAGGATTACTGATCTGTTTGGCGCATCCACACATGTGGAT CATGTGCTGCCCTGGTCTGTAACCTGCGACGACGGGCTGGCTA ACAAGGTGGTTTGTACGGCTACTGCCAATCTGTTCAAGGGCAA CAATACACCGCATGATGCTTTCTCCGGCGCAGAGGGCTATGAC TGGGCTGCAATCCTGCGTCGGGTAGACAGGATGATTCCAGCTA	MGGFFPYRLALDGGIASFG WAIYALDRNGEPCDLIDGN VLLFPTAIGASNRREKRSAR KTNDRKQLRLKKICEFFASD FQGYNREKAPADVLNTSPF RLRANALDRQVSLAELQRI CLHMAKHRGSSAIRMSEDE EIAKEGAAISSAFARTQVQM LERSARTYGELLWRIEKAN NRKSGQDTTKIRISKQGNK AGDAYSHYPLRFLIEQEFDK IMLAQSAFHDELTQEKINAL KEILFFERETITPKPGYCLYI ENERRIPKLSRLFQLKRIYEE VNNLRIDTAEGIRILTFIERD LLIDQLMAGVTLGKKNTKE FLKLKGKDKLYLGKKDTLV IKPYPFDCLLGQDAVLGDR WAEMSEQQQDELLDILATV WDDTRAHIALKKLLNCND VLARAILAVPLPSGYGAIGA SATREIIDVLEANICSATEAA MLAGLKERAIFEGDLLPRLP YYGEILGNHVSFSDNAGSSL PYQEKVFGRIPNMVVHRAL IQLRGLVNSIIKKYGPPTQIH LELGRELSMSEQERDDIQKT NENNRKVNDVIRNKLVKNE EKPNRSNLIRYKLWERQKG TCIYSGDNIRITDLFGASTHV DHVLPWSVTCDDGLANKV VCTATANLFKGNNTPHDAF SGAEGYDWAAILRRVDRMI PAMKWRFTADALEKFEDE NAGSFSARAMNDNRYIAKL MRFYLSSICRQQDIVVSNGR ITSNLNRAWGLDKFFVDHS EIIQEQIGKGKGRSKRADHR HHFVDAVAIGGVTRGMIQK IQTEAGRVEREGAENEFFRR VFAPLEPPVRDLEKRVLALL QQMRPVYKNNHKKQGQLH LETRNGIIDGPDRHGYYINR VRKKITDLPNLEKLNALRIE QTIPDLDEVVKSRERLKAIK

	TGAAGTGGCGGTTTACAGCCGATGCCTTGGAAAAATTTGAAGA TGAAAATGCCGGTTCATTTTCCGCTAGAGCCATGAATGACAAC CGCTATATCGCAAAGTTGATGCGGTTCTATTTATCGTCGATTTG CCGGCAGCAGGATATTGTCGTTTCAAATGGCCGTATTACCAGC AACTTGAACCGGGCCTGGGGTCTTGATAAATTTTTTGTCGATCA TTCCGAAATCATACAGGAACAGATTGGCAAGGGAAAAGGCCG ATCAAAACGGGCTGATCACCGGCATCATTTTGTTGATGCGGTTG CAATCGGCGGCGTCACCCGGGGAATGATCCAGAAAATTCAGAC CGAAGCCGGGCGCGTGGAAAGGGAGGGAGCCGAGAATGAATT TTTCCGTCGGGTCTTTGCGCCTCTGGAACCACCGGTTCGCGACC TTGAAAAAAGGGTTTTGGCTTTATTGCAACAGATGCGTCCTGTT TACAAAAATAACCACAAGAAACAGGGGCAATTGCATCTGGAA ACCCGTAATGGCATTATCGACGGGCCTGACAGGCACGGCTATT ATATCAATCGCGTTCGTAAAAAGATCACAGATCTTCCCAATCTT GAGAAGCTGAACGCGCTCAGAATTGAACAGACAATTCCGGATC TTGATGAAGTTGTCAAATCCAGAGAGCGTCTGAAGGCAATCAA AACAGGAATAGAAAACTGTTGGGAAGCGGCCCGAGCTAAATT ACTAAGCGAAAATCAACAGGCAATCGTCGAGAACAACAAACC CAAAGAAATCACCGATACGGTGGTTTTTAAAGTCGCGCGGGAC CTGTATCTTTCTGCGGGCGGTGCAAACGAATATGTCATTTATAA ACGCGAGAAACTGATCGTGCCCGCACATGCGATTGTGGGTAAT CGACCTCGGTTTGGTTATGTCTCGGGTAATAATCAACGGATTGA CTTTTACCGGAATCGCGGTACCGGAAAATTGGGCTGGCAGTGT ATCAGCCACTATGAAGCCATAAACGAAGACAGAAAAGATGCCT TGTTCATACCTGAGAGCAAAAAAGACGTCAATGAATTGATCTT CTCGCTGTTTAAGGATGACCTTGTAGAAATGGATTGTCCGATAG GTGAGGGGACGCGGGATCTTTATAAGGTCGCCAAATTTACCAA GGGTAAAATGATGGTCGTCCCCGTCCTTGATGCCCGCATGTCA ACCGATGGGCGTCAGGACCTTGGTGGCAGAGGTCTGGGGTTTT ATTGCGATGCCAAGGCACGTAAAATCAAACAAGATAAAACGG GAAAAATCATATGGAAATCAAGTCGCTTACAATAG	TGIENCWEAARAKLLSENQ QAIVENNKPKEITDTVVFKV ARDLYLSAGGANEYVIYKR EKLIVPAHAIVGNRPRFGYV SGNNQRIDFYRNRGTGKLG WQCISHYEAINEDRKDALFI PESKKDVNELIFSLFKDDLV EMDCPIGEGTRDLYKVAKF TKGKMMVVPVLDARMSTD GRQDLGGRGLGFYCDAKA RKIKQDKTGKIIWKSSRLQ*

sgRNA sequences used in chapter 3 with the following format:

Spacer – crRNA – Linker –tracrRNA

Cas9	sgRNA
NmeCas9	N24GTTGTAGCTCCCTTTCTCATTTCGGAAA CGAAATGAGAACCGTTGCTACAATAAGGCCG TCTGAAAAGATGTGCCGCAACGCTCTGCCCC TTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
BoeCas9	N24GTTGTAGTTCCCCATTCCACTCTGTAGTGCTATAATGAAA GTTGTAGCACTACAGAGTGGAATCCGTTACTACAATAAAAGAA TTACTTGTGAAGCCCTGCCCCTATTAGCATCTGCTATTAGGGG CATCTTCTTTTCTAC
KlaCas9	N24GTTACGGCTTCCCTGCTAATCGATGAAA ATCGATTAGCAGGGATCAAGAGTCGTAATAA GTATTTATTACGCAAAATGGGGTGCTTACGG GCACCCCTTCTTCGTTTGTAAATGGATGTGT TATTAGTCGCGCCTA
HpaCas9	N24GTTGTAGCTCCCTTTTTCATTTCGCAGAAA TGCGAAATGAAAAACGTTGTTACAATAAGAGAA AAGATTTCTCGCAAAGCTCTGTCCCTTGAAATG TAAGTTTCAAGGGACATCTTT
BdeCas9	N24GTTGTAGCTCCCTCTCTCATCTCGTAGTGGAAA CACTACGAGATGAGAGCCGTTGCTACAATAAGGCCG TCTGAAAAGACGCGCCGCGACGTAAAATACTTTATG TATGAGCCCCTGTTTGAGTTTTCTTAAACAGGGGCA TCGTCTT

Plasmids used in chapter 3:

Name	Insert description	Backbone	purpose
AE51	HpaCas9	pMSCG7	Bacterial expression
AE52	SmuCas9	pMSCG7	Bacterial expression
AE53	KlaCas9	pMSCG7	Bacterial expression
AE54	BoeCas9	pMSCG7	Bacterial expression
AE65	Geo_sgRNA	pLKO	sgRNA Cassette
AE55	Hpa_sgRNA	pLKO	sgRNA Cassette
AE56	Kla_sgRNA	pLKO	sgRNA Cassette
AE58	HpaCas9	pcDest	mammalian expression
AE59	SmuCas9	pcDest	mammalian expression
AE60	KlaCas9	pcDest	mammalian expression

Appendix 2: Information related to chapter 4

Sequences: SV40 NLS 3X-HA-Tag cMyc-like NLS

Native Nme2Cas9 bacterial DNA sequence:

ATGGCTGCCTTCAAACCTAATCCAATCAACTACATCCTCGGCCTCGATATCGGCATCGCATC
CGTCGGCTGGGCGATGGTAGAAATTGACGAAGAAGAAAACCCCATCCGCCTGATTGATTTG
GGCGTGCGCGTATTTGAGCGTGCCGAAGTACCGAAAACAGGCGACTCCCTTGCCATGGCAA
GGCGTTTGGCGCGCAGTGTCCGCCGCCTGACCCGCCGTCGCGCCCATCGCCTGCTTCGGGCC
CGCCGCCTATTGAAACGCGAAGGCGTATTACAAGCCGCTGATTTTGACGAAAACGGCTTGA
TCAAATCCTTACCGAATACACCATGGCAACTTCGCGCAGCCGCATTAGACCGCAAACTGAC
GCCTTTAGAGTGGTCGGCAGTCTTGTTGCATTTAATCAAACACCGCGGTTATTTGTCGCAAA
GAAAAAACGAGGGCGAAACTGCCGATAAAGAGCTTGGCGCTTTGCTTAAAGGCGTGGCCAA
CAATGCCCATGCCTTACAGACAGGCGATTTCCGCACACCGGCCGAATTGGCTTTAAATAAAT
TTGAGAAAGAAAGCGGCCATATCCGCAATCAGCGCGGCGATTATTCGCATACGTTCAGCCG
CAAAGATTTACAGGCGGAGCTGATTTTGCTGTTTGAAAAACAAAAAGAATTTGGCAATCCG
CATGTTTCAGGCGGCCTTAAAGAAGGTATTGAAACCCTACTGATGACGCAACGCCCTGCCCT
GTCCGGCGATGCCGTTCAAAAAATGTTGGGGCATTGCACCTTCGAACCGGCAGAGCCGAAA
GCCGCTAAAAACACCTACACAGCCGAACGTTTCATCTGGCTGACCAAGCTGAACAACCTGC
GTATTTTAGAGCAAGGCAGCGAGCGGCCATTGACCGATACCGAACGCGCCACGCTTATGGA
CGAGCCATACAGAAAATCCAAACTGACTTACGCACAAGCCCGTAAGCTGCTGGGTTTAGAA
GATACCGCCTTTTTCAAAGGCTTGCGCTATGGTAAAGACAATGCCGAAGCCTCAACATTGAT
GGAAATGAAGGCCTACCATGCCATCAGCCGTGCACTGGAAAAAGAAGGATTGAAAGATAA
AAAATCCCCATTAAACCTTTCTTCCGAATTACAAGATGAAATCGGCACGGCATTCTCCCTGT
TCAAAACCGATGAAGACATTACAGGCCGTCTGAAAGACCGTGTTCAGCCCGAAATCTTAGA
AGCGCTGTTGAAACACATCAGCTTCGATAAGTTCGTCCAAATTTCCTTGAAAGCATTGCGCC
GAATTGTGCCTCTAATGGAGCAAGGCAAACGTTACGATGAAGCCTGCGCCGAAATCTACGG
AGACCATTACGGCAAGAAAAATACGGAAGAAAAAATTTATCTGCCGCCCATCCCTGCCGAC
GAGATCCGCAACCCCGTCGTCTTGCGCGCCTTATCTCAAGCACGTAAGGTCATTAACGGCGT
GGTACGCCGTTACGGCTCCCCAGCTCGTATCCATATTGAAACGGCAAGGGAAGTAGGTAAA
TCGTTTAAAGACCGCAAAGAAATCGAAAAACGCCAAGAAGAAAACCGCAAAGACCGGGAA
AAAGCCGCCGCCAAATTCCGAGAGTATTTCCCCAATTTTGTCGGCGAACCCAAATCAAAAG
ATATTCTGAAACTGCGCCTGTACGAGCAACAACACGGCAAATGCCTGTATTCGGGCAAAGA
AATCAACTTAGTCCGTCTGAACGAAAAAGGCTATGTCGAAATCGACCATGCCCTGCCGTTTT
CGCGCACATGGGACGACAGTTTCAACAATAAAGTGCTGGTATTGGGCAGCGAAAACCAAAA
CAAAGGCAATCAAACCCCTTACGAATACTTCAACGGCAAAGACAACAGCCGCGAATGGCAG
GAATTTAAAGCGCGTGTCGAAACCAGCCGTTTCCCGCGCAGTAAAAAACAACGGATTCTGC
TGCAAAAATTCGATGAAGACGGCTTTAAAGAATGCAATCTGAACGACACGCGCTACGTCAA
CCGTTTCCTGTGCCAATTTGTTGCCGACCATATATTGCTGACAGGTAAAGGGAAAAGACGTG
TCTTTGCCTCAAACGGACAAATTACCAATCTGTTGCGCGGCTTTTGGGGATTGCGCAAAGTG
CGTGCGGAAAACGACCGCCATCACGCCTTGGACGCTGTAGTCGTTGCCTGCTCGACCGTTGC
CATGCAGCAGAAAATTACCCGTTTTGTACGCTATAAAGAGATGAACGCGTTTGACGGTAAA
ACCATAGACAAAGAAACAGGAAAAGTGCTGCATCAAAAAACACACTTCCCACAACCTTGGG
AATTTTTCGCACAAGAAGTCATGATTCGCGTCTTCGGCAAACCGGACGGCAAACCCGAATTC
GAAGAAGCCGATACCCCAGAAAAACTGCGCACGTTGCTTGCCGAAAAATTATCATCTCGCC
CCGAAGCCGTACACGAATACGTTACGCCACTGTTTGTTTCACGCGCGCCCAATCGGAAGATG
AGCGGTGCACATAAAGATACTTTGAGATCTGCTAAACGATTTGTTAAACATAATGAAAAAA
TTAGTGTTAAACGAGTATGGTTAACCGAAATCAAGTTGGCCGACCTTGAAAATATGGTTAAT
TATAAAAATGGTAGAGAGATTGAATTATATGAGGCTCTTAAGGCGCGTTTAGAGGCATATG
GAGGTAATGCTAAACAAGCATTTGACCCTAAGGACAATCCGTTTTATAAAAAGGGAGGACA
ACTGGTTAAAGCTGTGAGGGTCGAAAAAACCCAAGAGAGCGGAGTCTTATTAAATAAAAAA
AATGCTTATACCATTGCAGATAATGGAGACATGGTACGTGTTGATGTGTTCTGTAAAGTAGA
TAAGAAAGGAAAAAATCAGTATTTTATTGTTCCCATCTATGCTTGGCAGGTTGCTGAAAACA
TTCTTCCCGATATTGATTGTAAGGGATACCGGATTGATGATAGCTATACATTCTGTTTTAGCT

TGCATAAGTATGATCTGATTGCTTTTCAAAAAGATGAAAAATCTAAAGTAGAATTCGCCTAC
TATATCAACTGTGATAGCTCTAATGGACGATTCTATTTAGCTTGGCATGATAAAGGCTCTAA
AGAACAGCAATTCCGTATTAGCACCCAAAATCTTGTATTGATACAAAAATACCAAGTTAAC
GAACTGGGCAAAGAAATCAGACCATGCCGTCTGAAAAAACGCCCACCTGTCCGTTAA

Nme2Cas9 bacterial protein sequence:
MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLA
RSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWS
AVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGH
IRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKML
GHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQ
ARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIG
TAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIY
GDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFK
DRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRL
NEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVET
SRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASNGQITNLL
RGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLH
QKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSR
APNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMVNYKNGREIELYEALKAR
LEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKKNAYTIADNGDMVRVDV
FCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSYTFCFSLHKYDLIAFQKDEKSKVEF
AYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQKYQVNELGKEIRPCRLKKRPPVR

Nme2Cas9 human-codon-optimized DNA sequence:
ATGGCCGCCTTCAAGCCTAACCCAATCAATTACATCCTGGGACTGGACATCGGAATCGCATC
CGTGGGATGGGCTATGGTGGAGATCGACGAGGAGGAGAATCCTATCCGGCTGATCGATCTG
GGCGTGAGAGTGTTTGAGAGGGCCGAGGTGCCAAAGACCGGCGATTCTCTGGCTATGGCCC
GGAGACTGGCACGGAGCGTGAGGCGCCTGACACGGAGAAGGGCACACAGGCTGCTGAGGG
CACGCCGGCTGCTGAAGAGAGAGGGCGTGCTGCAGGCAGCAGACTTCGATGAGAATGGCCT
GATCAAGAGCCTGCCAAACACCCCCTGGCAGCTGAGAGCAGCCGCCCTGGACAGGAAGCTG
ACACCACTGGAGTGGTCTGCCGTGCTGCTGCACCTGATCAAGCACCGCGGCTACCTGAGCC
AGCGGAAGAACGAGGGAGAGACAGCAGACAAGGAGCTGGGCGCCCTGCTGAAGGGAGTGG
CCAACAATGCCCACGCCCTGCAGACCGGCGATTTCAGGACACCTGCCGAGCTGGCCCTGAA
TAAGTTTGAGAAGGAGTCCGGCCACATCAGAAACCAGAGGGGCGACTATAGCCACACCTTC
TCCCGCAAGGATCTGCAGGCCGAGCTGATCCTGCTGTTCGAGAAGCAGAAGGAGTTTGGCA
ATCCACACGTGAGCGGAGGCCTGAAGGAGGGAATCGAGACCCTGCTGATGACACAGAGGC
CTGCCCTGTCCGGCGACGCAGTGCAGAAGATGCTGGGACACTGCACCTTCGAGCCTGCAGA
GCCAAAGGCCGCCAAGAACACCTACACAGCCGAGCGGTTTATCTGGCTGACAAAGCTGAAC
AATCTGAGAATCCTGGAGCAGGGATCCGAGAGGCCACTGACCGACACAGAGAGGGCCACC
CTGATGGATGAGCCTTACCGGAAGTCTAAGCTGACATATGCCCAGGCCAGAAAGCTGCTGG
GCCTGGAGGACACCGCCTTCTTTAAGGGCCTGAGATACGGCAAGGATAATGCCGAGGCCTC
CACACTGATGGAGATGAAGGCCTATCACGCCATCTCTCGCGCCCTGGAGAAGGAGGGCCTG
AAGGACAAGAAGTCCCCCCTGAACCTGAGCTCCGAGCTGCAGGATGAGATCGGCACCGCCT
TCTCTCTGTTTAAGACCGACGAGGATATCACAGGCCGCCTGAAGGACAGGGTGCAGCCTGA
GATCCTGGAGGCCCTGCTGAAGCACATCTCTTTCGATAAGTTTGTGCAGATCAGCCTGAAGG
CCCTGAGAAGGATCGTGCCACTGATGGAGCAGGGCAAGCGGTACGACGAGGCCTGCGCCG
AGATCTACGGCGATCACTATGGCAAGAAGAACACAGAGGAGAAGATCTATCTGCCCCCTAT
CCCTGCCGACGAGATCAGAAATCCTGTGGTGCTGAGGGCCCTGTCCCAGGCAAGAAAAGTG
ATCAACGGAGTGGTGCGCCGGTACGGATCTCCAGCCCGGATCCACATCGAGACCGCCAGAG
AAGTGGGCAAGAGCTTCAAGGACCGGAAGGAGATCGAGAAGAGACAGGAGGAGAATCGCA
AGGATCGGGAGAAGGCCGCCGCCAAGTTTAGGGAGTACTTCCCTAACTTTGTGGGCGAGCC

AAAGTCTAAGGACATCCTGAAGCTGCGCCTGTACGAGCAGCAGCACGGCAAGTGTCTGTAT
AGCGGCAAGGAGATCAATCTGGTGCGGCTGAACGAGAAGGGCTATGTGGAGATCGATCAC
GCCCTGCCTTTCTCCAGAACCTGGGACGATTCTTTTAACAATAAGGTGCTGGTGCTGGGCAG
CGAGAACCAGAATAAGGGCAATCAGACACCATACGAGTATTTCAATGGCAAGGACAACTCC
AGGGAGTGGCAGGAGTTCAAGGCCCGCGTGGAGACCTCTAGATTTCCCAGGAGCAAGAAGC
AGCGGATCCTGCTGCAGAAGTTCGACGAGGATGGCTTTAAGGAGTGCAACCTGAATGACAC
CAGATACGTGAACCGGTTCCTGTGCCAGTTTGTGGCCGATCACATCCTGCTGACCGGCAAGG
GCAAGAGAAGGGTGTTCGCCTCTAATGGCCAGATCACAAACCTGCTGAGGGGATTTTGGGG
ACTGAGGAAGGTGCGGGCAGAGAATGACAGACACCACGCACTGGATGCAGTGGTGGTGGC
ATGCAGCACCGTGGCAATGCAGCAGAAGATCACAAGATTCGTGAGGTATAAGGAGATGAA
CGCCTTTGACGGCAAGACCATCGATAAGGAGACAGGCAAGGTGCTGCACCAGAAGACCCAC
TTCCCCCAGCCTTGGGAGTTCTTTGCCCAGGAAGTGATGATCCGGGTGTTCGGCAAGCCAGA
CGGCAAGCCTGAGTTTGAGGAGGCCGATACCCCAGAGAAGCTGAGGACACTGCTGGCAGA
GAAGCTGTCTAGCAGGCCAGAGGCAGTGCACGAGTACGTGACCCCACTGTTCGTGTCCAGG
GCACCCAATCGGAAGATGTCTGGCGCCCACAAGGACACACTGAGAAGCGCCAAGAGGTTTG
TGAAGCACAACGAGAAGATCTCCGTGAAGAGAGTGTGGCTGACCGAGATCAAGCTGGCCG
ATCTGGAGAACATGGTGAATTACAAGAACGGCAGGGAGATCGAGCTGTATGAGGCCCTGAA
GGCAAGGCTGGAGGCCTACGGAGGAAATGCCAAGCAGGCCTTCGACCCAAAGGATAACCC
CTTTTATAAGAAGGGAGGACAGCTGGTGAAGGCCGTGCGGGTGGAGAAGACCCAGGAGAG
CGGCGTGCTGCTGAATAAGAAGAACGCCTACACAATCGCCGACAATGGCGATATGGTGAGA
GTGGACGTGTTCTGTAAGGTGGATAAGAAGGGCAAGAATCAGTACTTTATCGTGCCTATCTA
TGCCTGGCAGGTGGCCGAGAACATCCTGCCAGACATCGATTGCAAGGGCTACAGAATCGAC
GATAGCTATACATTCTGTTTTTCCCTGCACAAGTATGACCTGATCGCCTTCCAGAAGGATGA
GAAGTCCAAGGTGGAGTTTGCCTACTATATCAATTGCGACTCCTCTAACGGCAGGTTCTACC
TGGCCTGGCACGATAAGGGCAGCAAGGAGCAGCAGTTTCGCATCTCCACCCAGAATCTGGT
GCTGATCCAGAAGTATCAGGTGAACGAGCTGGGCAAGGAGATCAGGCCATGTCGGCTGAAG
AAGCGCCCACCCGTGCGGGGCACCGGCGGGCCCAAGAAGAAGAGGAAGGTATACCCATAC
GATGTTCCTGACTATGCGGGCTATCCCTATGACGTCCCGGACTATGCAGGATCGTATCCTTA
TGACGTTCCAGATTACGCTGGATCCGCCGCTCCGGCAGCTAAGAAAAAGAAACTGGATTTC
GAATCCGGATAA

Nme2Cas9 humanized protein sequence:

MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLA
RSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWS
AVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGH
IRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKML
GHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQ
ARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIG
TAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIY
GDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFK
DRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRL
NEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVET
SRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASNGQITNLL
RGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLH
QKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSR
APNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMVNYKNGREIELYEALKAR
LEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKKNAYTIADNGDMVRVDV
FCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSYTFCFSLHKYDLIAFQKDEKSKVEF
AYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQKYQVNELGKEIRPCRLKKRPPVRGTG
GPKKKRKVYPYDVPDYAGYPYDVPDYAGSYPYDVPDYAGSAAPAAKKKKLDFESG*

Target sites used in chapter 4

#	Site Name	Spacer Seq	PAM	Locus	Editing (%)
1	TS1	GGTTCTGGGTACTTTTATCTGTCC	CCTCCACC	*AAVS1*	ND
2	TS4	GTCTGCCTAACAGGAGGTGGGGGT	TAGACGAA	*AAVS1*	11
3	TS5	GAATATCAGGAGACTAGGAAGGAG	GAGGCCTA	*AAVS1*	15
4	TS6	GCCTCCCTGCAGGGCTGCTCCC	CAGCCCAA	*LINC01588*	20
5	TS10	GAGCTAGTCTTCTTCCTCCAACCC	GGGCCCTA	*AAVS1*	3.5
6	TS11	GATCTGTCCCCTCCACCCCACAGT	GGGGCCAC	*AAVS1*	9
7	TS12	GGCCCAAATGAAAGGAGTGAGAGG	TGACCCGA	*AAVS1*	10
8	TS13	GCATCCTCTTGCTTTCTTTGCCTG	GACACCCC	*AAVS1*	2
9	TS16	GGAGTCGCCAGAGGCCGGTGGTGG	ATTTCCTC	*LINC01588*	28
10	TS17	GCCCAGCGGCCGGATATCAGCTGC	CACGCCCG	*LINC01588*	ND
11	TS18	GGAAGGGAACATATTACTATTGC	TTTCCCTC	*CYBB*	1
12	TS19	GTGGAGTGGCCTGCTATCAGCTAC	CTATCCAA	*CYBB*	6
13	TS20	GAGGAAGGGAACATATTACTATTG	CTTTCCCT	*CYBB*	11.2
14	TS21	GTGAATTCTCATCAGCTAAAATGC	CAAGCCTT	*CYBB*	1
15	TS25	GCTCACTCACCCACACAGACACAC	ACGTCCTC	*VEGF A*	15.6
16	TS26	GGAAGAATTTCATTCTGTTCTCAG	TTTTCCTG	*CFTR*	2
17	TS27	GCTCAGTTTTCCTGGATTATGCCT	GGCACCAT	*CFTR*	4
18	TS31	GCGTTGGAGCGGGGAGAAGGCCAG	GGGTCACT	*VEGFA*	9
19	TS34	GGGCCGCGGAGATAGCTGCAGGGC	GGGGCCCC	*LINC01588*	ND
20	TS35	GCCCACCCGGCGGCGCCTCCCTGC	AGGGCTGC	*LINC01588*	ND
21	TS36	GCGTGGCAGCTGATATCCGGCCGC	TGGGCGTC	*LINC01588*	ND
22	TS37	GCCGCGGCGCGACGTGGAGCCAGC	CCCGCAAA	*LINC01588*	ND
23	TS38	GTGCTCCCCAGCCCAAACCGCCGC	GGCGCGAC	*LINC01588*	2
24	TS41	GTCAGATTGGCTTGCTCGGAATTG	CCAGCCAA	*AGA*	3
25	TS44	GCTGGGTGAATGGAGCGAGCAGCG	TCTTCGAG	*VEGFA*	3
26	TS45	GTCCTGGAGTGACCCCTGGCCTTC	TCCCCGCT	*VEGFA*	7.4
27	TS46	GATCCTGGAGTGACCCCTGGCCTT	CTCCCCGC	*VEGFA*	6
28	TS47	GTGTGTCCCTCTCCCCACCCGTCC	CTGTCCGG	*VEGFA*	23.1
29	TS48	GTTGGAGCGGGGAGAAGGCCAGGG	GTCACTCC	*VEGFA*	2
30	TS49	GCGTTGGAGCGGGGAGAAGGCCAG	GGGTCACT	*VEGFA*	4
31	TS50	GTACCCTCCAATAATTTGGCTGGC	AATTCCGA	*AGA*	6
32	TS51	GATAATTTGGCTGGCAATTCCGAG	CAAGCCAA	*AGA*	4.5
33	TS58 (DS1)	GCAGGGGCCAGGTGTCCTTCTCTG	GGGGCCTC	*VEGFA*	5

34	TS59 (DS2)	GAATGGCAGGCGGAGGTTGTACTG	GGGGCCAG	*VEGFA*	11.5
35	TS60 (DS3)	GAGTGAGAGAGTGAGAGAGAGACA	CGGGCCAG	*VEGFA*	3
36	TS61 (DS4)	GTGAGCAGGCACCTGTGCCAACAT	GGGCCCGC	*VEGFA*	3.5
37	TS62 (DS5)	GCGTGGGGGCTCCGTGCCCCACGC	GGGTCCAT	*VEGFA*	3.4
38	TS63 (DS6)	GCATGGGCAGGGGCTGGGGTGCAC	AGGCCCAG	*VEGFA*	16
39	TS64	GAAAATTGTGATTTCCAGATCCAC	AAGCCCAA	*FANCJ*	7
40	TS65	GAGCAGAAAAAATTGTGATTTCC	AGATCCAC	*FANCJ*	ND
41	TS66 (DS7)	GGCCGCGGAGATAGCTGCAGGGCG	GGGCCCCC	*LINC01588*	1
42	TS67 (DS8)	GGAGGGGCTGGGGCCTGGACTCCT	GGGTCCGA	*AAVS1*	ND
43	TS68 (DS9)	GTCGGGCCCGCACCCGCGCCCGGC	AGGGCCAG	Intergenic	ND
44	TS69 (DS10)	GCCTGTCCCGCGCTCCCGCCCGGC	AGGCCCGG	Intergenic	3
45	TS70 (DS11)	GTCTCTCTCTCTCTCTCTCCAGGC	TGGTCCAT	Intergenic	3
46	TS71 (DS12)	GCGCAAAGCTGCATCCACCCCCCG	AGGACCAG	*LINC01588*	49
47	TS72 (DS13)	GACCAGCCCCTCGAAGGCAAGGCC	AGGACCTC	*LINC01588*	32
48	TS73 (DS14)	GGAGTTTTCTCTTAGGCTTCTCTC	AGGTCCCC	*LINC01588*	50
49	TS74 (DS15)	GGGGGCGCATGGCTCCGCCCCGCC	GGGACCCC	VEGFA	5
50	TS75 (DS16)	GTTCCCCTTCATTGCGGCGGGCTG	CGGGCCAG	VEGFA	35
51	TS76 (DS17)	CATGGAACAAACTGACTGAAGGCG	AGGTCCGG	*ARHGEF9*	5
52	TS77 (DS18)	CCTACAACCCAATCCCCTCCGCCC	CGGACCTC	*ARHGEF9*	6
53	TS78 (DS19)	CCCGCGACAACTCGCGCCAGCTAC	GGGGCCTC	*ARHGEF9*	ND
54	TS79 (DS20)	TGCTTGCGAAGTCCGGCTTCTCTG	AGGCCCCG	*ARHGEF9*	ND
55	TS80 (DS21)	GTGGGCTTGCAGGAGGCAGGTCTG	TGGCCCCT	*LSP1*	1
56	TS81 (DS22)	TATGTTCCAGCTTCCTGGGTCTGC	AGGTCCAG	*LSP1*	39
57	TS82 (DS23)	GATTCCAGCCAGACACCCGCCCCC	CGGCCCTG	*LSP1*	10
58	TS83 (DS24)	GCAACAGGAAGCAACTTCTTAGCC	AGGGCCGG	*LSP1*	ND
59	TS85 (DS25)	CATCCATCGGCGCTTTGGTCGGCA	TGGCCCCA	*FancF*	ND
60	TS86 (DS26)	CCGCCGCTCCAGAGCCGTGCGAAT	GGGGCCAT	*FancF*	ND
61	TS87 (DS27)	TCTCTCTGCGCCTGCTGGAGAACC	GGGCCCTC	*FancF*	ND
62	TS88 (DS28)	GAGGCAAGAGGGCGGCTTTGGGCG	GGGTCCAG	*FancF*	33
63	TS89	GCCCTCATCTCCAATATGGTATGG	CGGCCCTT	*SEC61B*	37
64	TS90	GGCAAGAGCACAAGAGGAAGAGAG	AGACCCTC	*GAPHD*	38
65	TS91	GTTGTTTCTCATTCCACATCATCT	TCAGCCAA	*TOMM20*	ND

#	Name	Sequence	Purpose
1	AAVS1_TIDE1_FW	TGGCTTAGCACCTCTCCAT	TIDE analysis
2	LINC01588_TIDE_FW	AGAGGAGCCTTCTGACTGCTGCAGA	TIDE analysis
3	AAVS1_TIDE2_FW	TCCGTCTTCCTCCACTCC	TIDE analysis
4	NTS55_TIDE_FW	TAGAGAACTGGGTAGTGTG	TIDE analysis
5	VEGF_TIDE3_FW	GTACATGAAGCAACTCCAGTCCCA	TIDE analysis
6	hCFTR_TIDE1_FW	TGGTGATTATGGGAGAACTGGAGC	TIDE analysis
7	AGA_TIDE1_FW	GGCATAAGGAAATCGAAGGTC	TIDE analysis
8	VEGF_TIDE4_FW	ACACGGGCAGCATGGGAATAGTC	TIDE analysis
9	VEGF_TIDE5_FW	CCTGTGTGGCTTTGCTTTGGTCG	TIDE analysis
10	VEGF_TIDE6_FW	GGAGGAAGAGTAGCTCGCCGAGG	TIDE analysis
11	VEGF_TIDE7_FW	AGGGAGAGGGAAGTGTGGGGAAGG	TIDE analysis
12	AAVS1_TIDE1_RV	AGAACTCAGGACCAACTTATTCTG	TIDE analysis
13	LINC01588_TIDE_RV	ATGACAGACACAACCAGAGGGCA	TIDE analysis
14	AAVS1_TIDE2_RV	TAGGAAGGAGGAGGCCTAAG	TIDE analysis
15	NTS55_TIDE_RV	CCAATATTGCATGGGATGG	TIDE analysis
16	VEGF_TIDE3_RV	ATCAAATTCCAGCACCGAGCGC	TIDE analysis
17	hCFTR_TIDE1_RV	ACCATTGAGGACGTTTGTCTCAC	TIDE analysis
18	AGA_TIDE1_RV	CATGTCCTCAAGTCAAGAACAAG	TIDE analysis
19	VEGF_TIDE4_RV	GCTAGGGGAGAGTCCCACTGTCCA	TIDE analysis
20	VEGF_TIDE5_RV	GTAGGGTGTGATGGGAGGCTAAGC	TIDE analysis
21	VEGF_TIDE6_RV	AGACCGAGTGGCAGTGACAGCAAG	TIDE analysis
22	VEGF_TIDE7_RV	GTCTTCCTGCTCTGTGCGCACGAC	TIDE analysis
23	RandomPAM_FW	TAGCGGCCGCTCATGCGCGGCGCATT ACCTTTACNNNNNNNNNNGGAT CCTCTAGAGTCG	Protospacer with randomized PAM
24	RandomPAM_RV	ACAGGAAACAGCTATGACCATGAAAGC TTGCATGCCTGCAGGTCGACTCTA GAGGATC	Protospacer with randomized PAM
25	DS2_ON_FW1	ctacacgacgctcttccgatctCCTGGAGCGTGTACG TTGG	Targeted Deep Seq
26	SpyDS2_OT1_FW1	ctacacgacgctcttccgatctCCTGTGGTCCCAGCT ACTTG	Targeted Deep Seq
27	SpyDS2_OT2_FW1	ctacacgacgctcttccgatctATCTGCGATGTCCTC GAGG	Targeted Deep Seq

28	SpyDS2_OT3_FW1	ctacacgacgctcttccgatctTGGTGTGCGCCTCTAACG	Targeted Deep Seq
29	SpyDS2_OT4_FW1	ctacacgacgctcttccgatctGGAGTCTTGCTTTGTCACTCAGA	Targeted Deep Seq
30	SpyDS2_OT5_FW1	ctacacgacgctcttccgatctAGCCTAGACCCAGTCCCAT	Targeted Deep Seq
31	SpyDS2_OT6_FW1	ctacacgacgctcttccgatctGCTGGGCATAGTAGTGGACT	Targeted Deep Seq
32	SpyDS2_OT7_FW1	ctacacgacgctcttccgatctTGGGGAGGCTGAGACACGA	Targeted Deep Seq
33	SpyDS2_OT8_FW1	ctacacgacgctcttccgatctCTTGGGAGGCTGAGGCAAG	Targeted Deep Seq
34	DS2_ON_RV1	agacgtgtgctcttccgatctCAGGAGGATGAGAGCCAGG	Targeted Deep Seq
35	SpyDS2_OT1_RV1	agacgtgtgctcttccgatctCAGGGTCTCACTCTATCACCCA	Targeted Deep Seq
36	SpyDS2_OT2_RV1	agacgtgtgctcttccgatctACTGAATGGGTTGAACTTGGC	Targeted Deep Seq
37	SpyDS2_OT3_RV1	agacgtgtgctcttccgatctGAGACAGAATCTTGCTCTGTCTCC	Targeted Deep Seq
38	SpyDS2_OT4_RV1	agacgtgtgctcttccgatctTCCCAGCTACTTGGGAGGC	Targeted Deep Seq
39	SpyDS2_OT5_RV1	agacgtgtgctcttccgatctCCTGCCCAAATAGGGAAGCAG	Targeted Deep Seq
40	SpyDS2_OT6_RV1	agacgtgtgctcttccgatctTGGCGCCTTAGTCTCTGCTAC	Targeted Deep Seq
41	SpyDS2_OT7_RV1	agacgtgtgctcttccgatctGCATGAGACACAGTTTCACTCTG	Targeted Deep Seq
42	SpyDS2_OT8_RV1	agacgtgtgctcttccgatctGAGAGAGTCTCACTGCGTTGC	Targeted Deep Seq
43	DS4_ON_FW3	ctacacgacgctcttccgatctTCTCTCACCCACTGGGCAC	Targeted Deep Seq
44	DS4_ON_RV3	agacgtgtgctcttccgatctGCTTCCAGACGAGTGCAGA	Targeted Deep Seq
45	SpyDS4_OT1_FW1	ctacacgacgctcttccgatctAAGTTTTCAAACCAGAA GAACTACGAC	Targeted Deep Seq
46	SpyDS4_OT2_FW1	ctacacgacgctcttccgatctCCGGTATAAGTCCTGGAGCG	Targeted Deep Seq
47	SpyDS4_OT3_FW1	ctacacgacgctcttccgatctGCCAGGGAGCAATGGCAG	Targeted Deep Seq
48	SpyDS4_OT6_FW1	ctacacgacgctcttccgatctCCTCGAATTCCACGGGGTT	Targeted Deep Seq
49	DS16_ON_FW1	ctacacgacgctcttccgatctGTTGGTGGGAGGGAAGTGAG	Targeted Deep Seq
50	SpyDS6_OT1_FW1	ctacacgacgctcttccgatctGATGGCGGTTGTAGCGGC	Targeted Deep Seq
51	SpyDS6_OT2_FW1	ctacacgacgctcttccgatctCACATAAACCTATGTTTCAGCAGA	Targeted Deep Seq
52	SpyDS6_OT3_FW1	ctacacgacgctcttccgatctGCTAGTTGGATTGAAGCAGGGT	Targeted Deep Seq
53	SpyDS6_OT4_FW1	ctacacgacgctcttccgatctTTGAGTGCGGCAGCTTCC	Targeted Deep Seq
54	SpyDS6_OT6_FW1	ctacacgacgctcttccgatctATAACCCTCCCAGGCAAAGTC	Targeted Deep Seq
55	SpyDS6_OT7_FW1	ctacacgacgctcttccgatctAGCCTGCACATCTGAGCTC	Targeted Deep Seq
56	SpyDS6_OT8_FW1	ctacacgacgctcttccgatctGGAGCATTGAAGTGCCTGG	Targeted Deep Seq
57	DeDS6_ON_RV1	agacgtgtgctcttccgatctCAGCCTGGGACCACTGA	Targeted Deep Seq
58	SpyDS6_OT1_RV1	agacgtgtgctcttccgatctCATCCTCGACAGTCGCGG	Targeted Deep Seq
59	SpyDS6_OT2_RV1	agacgtgtgctcttccgatctGACTGATCAAGTAGAATACTCATGGG	Targeted Deep Seq

60	SpyDS6_OT3_RV1	agacgtgtgctcttccgatctCCCTGCCAGCACTGAAGC	Targeted Deep Seq
61	SpyDS6_OT4_Rv1	agacgtgtgctcttccgatctGGTTCCTATCTTTCTAGACCAGGAGT	Targeted Deep Seq
62	SpyDS6_OT6_RV1	agacgtgtgctcttccgatctAGTGTGGAGGGCTCAGGG	Targeted Deep Seq
63	SpyDS6_OT7_RV1	agacgtgtgctcttccgatctGATGGGCAGAGGAAGGCAA	Targeted Deep Seq
64	SpyDS6_OT8_RV1	agacgtgtgctcttccgatctTCACTCTCATGAGCGTCCCA	Targeted Deep Seq
65	Nme2DS2_OT1_FW1	ctacacgacgctcttccgatctAAGGTTCCTTGCGGTTCGC	Targeted Deep Seq
66	Nme2DS2_OT1_RV1	agacgtgtgctcttccgatctCGCTGCCATTGCTCCCT	Targeted Deep Seq
67	Nme2DS6_OT1_FW1	ctacacgacgctcttccgatctTCTCGCACATTCTTCACGTCC	Targeted Deep Seq
68	Nme2DS6_OT1_RV1	agacgtgtgctcttccgatctAGGAACCTTCCCGACTTAGGG	Targeted Deep Seq
69	Rosa26_ON_FW1	ctacacgacgctcttccgatctCCCGCCCATCTTCTAGAAAGAC	Targeted Deep Seq
70	Rosa26_OT1_FW1	ctacacgacgctcttccgatctTGCCAGGTGAGGGACTGG	Targeted Deep Seq
71	Rosa26_ON_RV1	agacgtgtgctcttccgatctTCTGGGAGTTCTCTGCTGCC	Targeted Deep Seq
72	Rosa26_OT1_RV1	agacgtgtgctcttccgatctTGCCCAACCTTAGCAAGGAG	Targeted Deep Seq
73	pCSK9_ON_FW2	ctacacgacgctcttccgatcttaccttggagcaacggcg	Targeted Deep Seq
74	PCSK9_ON_RV2	agacgtgtgctcttccgatctcccaggacgaggatggag	Targeted Deep Seq
75	Tyr_500_FW3	GATAGTCACTCCAGGGGTTG	TIDE analysis
76	Tyr_500_RV3	GTGGTGAACCAATCAGTCCT	TIDE analysis
77	Sec61b_TIDE_FW	CCGGAGGTCTCTACTTCCCAT	TIDE analysis
78	Sec61b_TIDE_RV	AAATGTCGCCTTCCTCCCAA	TIDE analysis
79	Gapdh_TIDE_FW	TAAAAAGTGCAGGGTCTGGCG	TIDE analysis
80	Gapdh_TIDE_RV	CTAACCAGTCAGCGTCAGAGC	TIDE analysis
81	Tomm20_TIDE_FW	GGATGCCGACTGTGTGTAAAG	TIDE analysis
82	Tomm20_TIDE_RV	CAGCAGCACACCTCACATAT	TIDE analysis
83	FANCF_TIDE_FW	CCAATCAGTACGCAGAGAGTC	TIDE analysis
84	FANCF_TIDE_RV	CCCAGAAGCCAGTGGACTAG	TIDE analysis
85	DTS3_TIDE_FW	GCCCCACAATAAACAGCATC	TIDE analysis
86	DTS3_TIDE_RV	CCCACCCCTATTCCACCTTA	TIDE analysis
87	DTS7_TIDE_FW	CTCCACCAGGAGTCAGGTCT	TIDE analysis
88	DTS7_TIDE_RV	CAGGGAGCTTTATTATACAAAATGG	TIDE analysis

Plasmids used in chapter 4

Name	Insert description	Backbone	purpose
pAE70	Nme3Cas9 PID on Nme1Cas9	pMCSG7	Bacterial expression of Nme1Cas9 with Nme3Cas9 PID
pAE71	Nme2Cas9 PID on Nme1Cas9	pMCSG7	Bacterial expression of Nme1Cas9 with Nme2Cas9 PID
pAE113	Nme2TLR1	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE114	Nme2TLR2	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE115	Nme2TLR5	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE116	Nme2TLR11	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE117	Nme2TLR12	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE118	Nme2TLR13	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE119	Nme2TLR14	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE120	Nme2TLR15	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE121	Nme2TLR16	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE122	Nme2TLR17	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE123	Nme2TLR18	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE124	Nme2TLR19	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE125	Nme2TLR20	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE126	Nme2TLR21	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE149	Nme2TLR22	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE150	Nme2TLR23	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE193	Nme2TLR13 with 23nt spacer	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE194	Nme2TLR13 with 22nt spacer	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE195	Nme2TLR13 with 21nt spacer	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE196	Nme2TLR13 with 20nt spacer	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE197	Nme2TLR13 with 19nt spacer	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE213	Nme2TLR21 with G22 spacer	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE214	Nme2TLR21 with G21 spacer	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE215	Nme2TLR21 with G20 spacer	pLKO	Targeting TLR2.0 with Nme2Cas9
pAE216	Nme2TLR21 with G19 spacer	pLKO	Targeting TLR2.0 with Nme2Cas9

pAE90	Nme2TS1	pLKO	Targeting *AAVS1* with Nme2Cas9
pAE93	Nme2TS4	pLKO	Targeting *AAVS1* with Nme2Cas9
pAE94	Nme2TS5	pLKO	Targeting *AAVS1* with Nme2Cas9
pAE129	Nme2TS6	pLKO	Targeting *LINC01588* with Nme2Cas9
pAE130	Nme2TS10	pLKO	Targeting *AAVS1* with Nme2Cas9
pAE131	Nme2TS11	pLKO	Targeting *AAVS1* with Nme2Cas9
pAE132	Nme2TS12	pLKO	Targeting *AAVS1* with Nme2Cas9
pAE133	Nme2TS13	pLKO	Targeting *AAVS1* with Nme2Cas9
pAE136	Nme2TS16	pLKO	Targeting *LINC01588* with Nme2Cas9
pAE137	Nme2TS17	pLKO	Targeting *LINC01588* with Nme2Cas9
pAE138	Nme2TS18	pLKO	Targeting *CYBB* with Nme2Cas9
pAE139	Nme2TS19	pLKO	Targeting *CYBB* with Nme2Cas9
pAE140	Nme2TS20	pLKO	Targeting *CYBB* with Nme2Cas9
pAE141	Nme2TS21	pLKO	Targeting *CYBB* with Nme2Cas9
pAE144	Nme2TS25	pLKO	Targeting *VEGFA* with Nme2Cas9
pAE145	Nme2TS26	pLKO	Targeting *CFTR* with Nme2Cas9
pAE146	Nme2TS27	pLKO	Targeting *CFTR* with Nme2Cas9
pAE152	Nme2TS31	pLKO	Targeting *VEGFA* with Nme2Cas9
pAE153	Nme2TS34	pLKO	Targeting *LINC01588* with Nme2Cas9
pAE154	Nme2TS35	pLKO	Targeting *LINC01588* with Nme2Cas9
pAE155	Nme2TS36	pLKO	Targeting *LINC01588* with Nme2Cas9
pAE156	Nme2TS37	pLKO	Targeting *LINC01588* with Nme2Cas9
pAE157	Nme2TS38	pLKO	Targeting *LINC01588* with Nme2Cas9
pAE159	Nme2TS41	pLKO	Targeting *AGA* with Nme2Cas9
pAE185	Nme2TS44	pLKO	Targeting *VEGFA* with Nme2Cas9
pAE186	Nme2TS45	pLKO	Targeting *VEGFA* with Nme2Cas9
pAE187	Nme2TS46	pLKO	Targeting *VEGFA* with Nme2Cas9
pAE188	Nme2TS47	pLKO	Targeting *VEGFA* with Nme2Cas9
pAE189	Nme2TS48	pLKO	Targeting *VEGFA* with Nme2Cas9
pAE190	Nme2TS49	pLKO	Targeting *VEGFA* with Nme2Cas9
pAE191	Nme2TS50	pLKO	Targeting *AGA* with Nme2Cas9
pAE192	Nme2TS51	pLKO	Targeting *AGA* with Nme2Cas9
pAE232	TS64_FancJ1	pLKO	Targeting *FANCJ* with Nme2Cas9
pAE233	TS65_FancJ2	pLKO	Targeting *FANCJ* with Nme2Cas9
pAE200	Nme2TS58 (Nme2DS1)	pLKO	Targeting dual site with Nme2Cas9

pAE201	Nme2TS59 (Nme2DS2)	pLKO	Targeting dual site with Nme2Cas9
pAE202	Nme2TS60 (Nme2DS3)	pLKO	Targeting dual site with Nme2Cas9
pAE203	Nme2TS61 (Nme2DS4)	pLKO	Targeting dual site with Nme2Cas9
pAE204	Nme2TS62 (Nme2DS5)	pLKO	Targeting dual site with Nme2Cas9
pAE205	Nme2TS63 (Nme2DS6)	pLKO	Targeting dual site with Nme2Cas9
pAE207	SpyDS1	pLKO	Targeting dual site with SpyCas9
pAE208	SpyDS2	pLKO	Targeting dual site with SpyCas9
pAE209	SpyDS3	pLKO	Targeting dual site with SpyCas9
pAE210	SpyDS4	pLKO	Targeting dual site with SpyCas9
pAE211	SpyDS5	pLKO	Targeting dual site with SpyCas9
pAE212	SpyDS6	pLKO	Targeting dual site with SpyCas9
pAE169	hDeCas9 Wt in AAV backbone	AAV	Nme2Cas9 all-in-one AAV expression with sgRNA cassette
pAE217	hDeCas9 wt in pMSCG7 backbone	pMCSG7	wildtype Nme2Cas9 for bacterial expression
pAE107	2xNLS Nme2Cas9 with HA	pCdest	Nme2Cas9 CMV-driven expression plasmid
pAE127	hDemonCas9 3X NLS in pMSCG7	pMSCG7	Targeting endogenous loci with Nme2Cas9
pAM172	hNme2Cas9 4X NLS with 3XHA	pCVL	Lentivector containing UCOE, SFFV driven Nme2Cas9 and Puro
pAM174	nickase hNme2Cas9 D16A 4X NLS with 3XHA	pCVL	Lentivector containing UCOE, SFFV driven Nme2Cas9 and Puro
pAM175	nickase hNme2Cas9 H588A 4X NLS with 3XHA	pCVL	Lentivector containing UCOE, SFFV driven Nme2Cas9 and Puro
pAM177	dead hNme2Cas9 4X NLS with 3XHA	pCVL	Lentivector containing UCOE, SFFV driven Nme2Cas9 and Puro
pAE170	TS66 (DS7)	pLKO	Targeting dual site with Nme2Cas9
pAE175	TS67 (DS8)	pLKO	Targeting dual site with Nme2Cas9
pAE257	TS68 (DS9)	pLKO	Targeting dual site with Nme2Cas9
pAE258	TS69 (DS10)	pLKO	Targeting dual site with Nme2Cas9
pAE262	TS70 (DS11)	pLKO	Targeting dual site with Nme2Cas9
pAE285	TS71 (DS12)	pLKO	Targeting dual site with Nme2Cas9
pAE286	TS72 (DS13)	pLKO	Targeting dual site with Nme2Cas9
pAE287	TS73 (DS14)	pLKO	Targeting dual site with Nme2Cas9
pAE288	TS74 (DS15)	pLKO	Targeting dual site with Nme2Cas9

pAE291	TS75 (DS16)	pLKO	Targeting dual site with Nme2Cas9
pAE299	TS76 (DS17)	pLKO	Targeting dual site with Nme2Cas9
pAE300	TS77 (DS18)	pLKO	Targeting dual site with Nme2Cas9
pAE301	TS78 (DS19)	pLKO	Targeting dual site with Nme2Cas9
pAE302	TS79 (DS20)	pLKO	Targeting dual site with Nme2Cas9
pAE303	TS80 (DS21)	pLKO	Targeting dual site with Nme2Cas9
pAE304	TS81 (DS22)	pLKO	Targeting dual site with Nme2Cas9
pAE305	TS82 (DS23)	pLKO	Targeting dual site with Nme2Cas9
pAE306	TS83 (DS24)	pLKO	Targeting dual site with Nme2Cas9
pAE307	TS85 (DS25)	pLKO	Targeting dual site with Nme2Cas9
pAE308	TS86 (DS26)	pLKO	Targeting dual site with Nme2Cas9
pAE309	TS87 (DS27)	pLKO	Targeting dual site with Nme2Cas9
pAE310	TS88 (DS28)	pLKO	Targeting dual site with Nme2Cas9
pAE178	SpyDS7	pLKO	Targeting dual site with SpyCas9
pAE181	SpyDS8	pLKO	Targeting dual site with SpyCas9
pAE263	SpyDS9	pLKO	Targeting dual site with SpyCas9
pAE264	SpyDS10	pLKO	Targeting dual site with SpyCas9
pAE268	SpyDS11	pLKO	Targeting dual site with SpyCas9
pAE312	SpyDS12	pLKO	Targeting dual site with SpyCas9
pAE313	SpyDS13	pLKO	Targeting dual site with SpyCas9
pAE316	SpyDS14	pLKO	Targeting dual site with SpyCas9
pAE318	SpyDS15	pLKO	Targeting dual site with SpyCas9
pAE320	SpyDS16	pLKO	Targeting dual site with SpyCas9
pAE325	SpyDS17	pLKO	Targeting dual site with SpyCas9
pAE326	SpyDS18	pLKO	Targeting dual site with SpyCas9
pAE327	SpyDS19	pLKO	Targeting dual site with SpyCas9
pAE328	SpyDS20	pLKO	Targeting dual site with SpyCas9
pAE329	SpyDS21	pLKO	Targeting dual site with SpyCas9
pAE330	SpyDS22	pLKO	Targeting dual site with SpyCas9
pAE331	SpyDS23	pLKO	Targeting dual site with SpyCas9
pAE332	SpyDS24	pLKO	Targeting dual site with SpyCas9
pAE333	SpyDS25	pLKO	Targeting dual site with SpyCas9
pAE334	SpyDS26	pLKO	Targeting dual site with SpyCas9
pAE335	SpyDS27	pLKO	Targeting dual site with SpyCas9
pAE336	SpyDS28	pLKO	Targeting dual site with SpyCas9

REFERENCES

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., Verdine, V., Cox, D. B. T., Kellner, M. J., Regev, A., Lander, E. S., Voytas, D. F., Ting, A. Y. and Zhang, F. (2017) 'RNA targeting with CRISPR-Cas13', *Nature,* 550(7675), pp. 280-284.

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V. and Zhang, F. (2016) 'C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector', *Science,* 353(6299), pp. aaf5573.

Agudelo, D., Carter, S., Velimirovic, M., Duringer, A., Levesque, S., Rivest, J.-F., Loehr, J., Mouchiroud, M., Cyr, D., Waters, P. J., Laplante, M., Moineau, S., Goulet, A. and Doyon, Y. (2019) 'Versatile and robust genome editing with Streptococcus thermophilus CRISPR1-Cas9', *bioRxiv*, pp. 321208.

Amrani, N., Gao, X. D., Liu, P., Edraki, A., Mir, A., Ibraheim, R., Gupta, A., Sasaki, K. E., Wu, T., Donohoue, P. D., Settle, A. H., Lied, A. M., McGovern, K., Fuller, C. K., Cameron, P., Fazzio, T. G., Zhu, L. J., Wolfe, S. A. and Sontheimer, E. J. (2018a) 'NmeCas9 is an intrinsically high-fidelity genome editing platform', *bioRxiv*, pp. https://doi.org/10.1101/172650.

Amrani, N., Gao, X. D., Liu, P., Edraki, A., Mir, A., Ibraheim, R., Gupta, A., Sasaki, K. E., Wu, T., Donohoue, P. D., Settle, A. H., Lied, A. M., McGovern, K., Fuller, C. K., Cameron, P., Fazzio, T. G., Zhu, L. J., Wolfe, S. A. and Sontheimer, E. J. (2018b) 'NmeCas9 is an intrinsically high-fidelity genome-editing platform', *Genome Biol,* 19(1), pp. 214.

Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., 3rd, Rademakers, R., Boylan, K. B., Dickson, D. W. and Petrucelli, L. (2013) 'Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS', *Neuron,* 77(4), pp. 639-46.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. and Horvath, P. (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes', *Science,* 315(5819), pp. 1709-12.

Basgall, E. M., Goetting, S. C., Goeckel, M. E., Giersch, R. M., Roggenkamp, E., Schrock, M. N., Halloran, M. and Finnigan, G. C. (2018) 'Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae', *Microbiology,* 164(4), pp. 464-474.

Berns, K. I. and Muzyczka, N. (2017) 'AAV: An Overview of Unanswered Questions', *Hum Gene Ther,* 28(4), pp. 308-313.

Bolukbasi, M. F., Gupta, A., Oikemus, S., Derr, A. G., Garber, M., Brodsky, M. H., Zhu, L. J. and Wolfe, S. A. (2015) 'DNA-binding-domain fusions enhance the targeting range and precision of Cas9', *Nat. Methods,* 12(12), pp. 1150-6.

Bondy-Denomy, J., Davidson, A. R., Doudna, J. A., Fineran, P. C., Maxwell, K. L., Moineau, S., Peng, X., Sontheimer, E. J. and Wiedenheft, B. (2018) 'A Unified Resource for Tracking Anti-CRISPR Names', *CRISPR J,* 1, pp. 304-305.

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. and Davidson, A. R. (2013) 'Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system', *Nature,* 493(7432), pp. 429-32.

Briner, A. E., Donohoue, P. D., Gomaa, A. A., Selle, K., Slorach, E. M., Nye, C. H., Haurwitz, R. E., Beisel, C. L., May, A. P. and Barrangou, R. (2014) 'Guide RNA functional modules direct Cas9 activity and orthogonality', *Mol Cell,* 56(2), pp. 333-339.

Brinkman, E. K., Chen, T., Amendola, M. and van Steensel, B. (2014) 'Easy quantitative assessment of genome editing by sequence trace decomposition', *Nucleic Acids Res.,* 42(22), pp. e168.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V. and van der Oost, J. (2008) 'Small CRISPR RNAs guide antiviral defense in prokaryotes', *Science,* 321(5891), pp. 960-4.

Burstein, D., Harrington, L. B., Strutt, S. C., Probst, A. J., Anantharaman, K., Thomas, B. C., Doudna, J. A. and Banfield, J. F. (2017) 'New CRISPR-Cas systems from uncultivated microbes', *Nature,* 542(7640), pp. 237-241.

Certo, M. T., Ryu, B. Y., Annis, J. E., Garibov, M., Jarjour, J., Rawlings, D. J. and Scharenberg, A. M. (2011) 'Tracking genome engineering outcome at individual DNA breakpoints', *Nat Methods,* 8(8), pp. 671-6.

Charlesworth, C. T., Deshpande, P. S., Dever, D. P., Camarena, J., Lemgart, V. T., Cromer, M. K., Vakulskas, C. A., Collingwood, M. A., Zhang, L., Bode, N. M., Behlke, M. A., Dejene, B., Cieniewicz, B., Romano, R., Lesch, B. J., Gomez-Ospina, N., Mantri, S., Pavel-Dinu, M., Weinberg, K. I. and Porteus, M. H. (2019) 'Identification of preexisting adaptive immunity to Cas9 proteins in humans', *Nat Med,* 25(2), pp. 249-254.

Charpentier, E., Richter, H., van der Oost, J. and White, M. F. (2015) 'Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity', *FEMS Microbiol Rev,* 39(3), pp. 428-41.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S. and Huang, B. (2013) 'Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system', *Cell,* 155(7), pp. 1479-91.

Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, J. K., Yildiz, A. and Doudna, J. A. (2017) 'Enhanced proofreading governs CRISPR-Cas9 targeting accuracy', *Nature*.

Cho, S. W., Kim, S., Kim, J. M. and Kim, J. S. (2013) 'Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease', *Nat Biotechnol,* 31(3), pp. 230-2.

Chylinski, K., Makarova, K. S., Charpentier, E. and Koonin, E. V. (2014) 'Classification and evolution of type II CRISPR-Cas systems', *Nucleic Acids Res,* 42(10), pp. 6091-105.

Cong, L. and Zhang, F. (2015) 'Genome engineering using CRISPR-Cas9 system', *Methods Mol Biol,* 1239, pp. 197-217.
Cyranoski, D. (2016) 'CRISPR gene-editing tested in a person for the first time', *Nature,* 539(7630), pp. 479.
De Caneva, A., Porro, F., Bortolussi, G., Sola, R., Lisjak, M., Barzel, A., Giacca, M., Kay, M. A., Vlahovicek, K., Zentilin, L. and Muro, A. F. (2019) 'Coupling AAV-mediated promoterless gene targeting to SaCas9 nuclease to efficiently correct liver metabolic diseases', *JCI Insight,* 5.
Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J. and Charpentier, E. (2011) 'CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III', *Nature,* 471(7340), pp. 602-7.
Ding, Y., Li, H., Chen, L. L. and Xie, K. (2016) 'Recent Advances in Genome Editing Using CRISPR/Cas9', *Front Plant Sci,* 7, pp. 703.
Dominguez, A. A., Lim, W. A. and Qi, L. S. (2016) 'Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation', *Nat Rev Mol Cell Biol,* 17(1), pp. 5-15.
Dong, Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., Yang, J., Xu, Z. and Huang, Z. (2017) 'Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein', *Nature,* 546(7658), pp. 436-439.
Doudna, J. A. and Charpentier, E. (2014) 'Genome editing. The new frontier of genome engineering with CRISPR-Cas9', *Science,* 346(6213), pp. 1258096.
East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H., Tjian, R. and Doudna, J. A. (2016) 'Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection', *Nature,* 538(7624), pp. 270-273.
Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J. and Church, G. M. (2013) 'Orthogonal Cas9 proteins for RNA-guided gene regulation and editing', *Nat Methods,* 10(11), pp. 1116-21.
Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K. S., Lecrivain, A. L., Bzdrenga, J., Koonin, E. V. and Charpentier, E. (2014) 'Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems', *Nucleic Acids Res,* 42(4), pp. 2577-90.
Franzosa, E. A. and Xia, Y. (2011) 'Structural principles within the human-virus protein-protein interaction network', *Proc Natl Acad Sci U S A,* 108(26), pp. 10538-43.
Freibaum, B. D. and Taylor, J. P. (2017) 'The Role of Dipeptide Repeats in C9ORF72-Related ALS-FTD', *Front Mol Neurosci,* 10, pp. 35.
Friedrich, G. and Soriano, P. (1991) 'Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice', *Genes Dev,* 5(9), pp. 1513-23.
Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. and Joung, J. K. (2014) 'Improving CRISPR-Cas nuclease specificity using truncated guide RNAs', *Nat Biotechnol,* 32(3), pp. 279-284.
Gallagher, D. N. and Haber, J. E. (2018) 'Repair of a Site-Specific DNA Cleavage: Old-School Lessons for Cas9-Mediated Gene Editing', *ACS Chem Biol,* 13(2), pp. 397-405.

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E. and James, A. A. (2015) 'Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi', *Proc Natl Acad Sci U S A,* 112(49), pp. E6736-43.

Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) 'Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria', *Proc Natl Acad Sci U S A,* 109(39), pp. E2579-86.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I. and Liu, D. R. (2017) 'Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage', *Nature,* 551(7681), pp. 464-471.

Germini, D., Tsfasman, T., Zakharova, V. V., Sjakste, N., Lipinski, M. and Vassetzky, Y. (2018) 'A Comparison of Techniques to Evaluate the Effectiveness of Genome Editing', *Trends Biotechnol,* 36(2), pp. 147-159.

Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J. and Zhang, F. (2018) 'Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6', *Science,* 360(6387), pp. 439-444.

Gorski, S. A., Vogel, J. and Doudna, J. A. (2017) 'RNA-based recognition and targeting: sowing the seeds of specificity', *Nat Rev Mol Cell Biol,* 18(4), pp. 215-228.

Grissa, I., Vergnaud, G. and Pourcel, C. (2007) 'CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats', *Nucleic Acids Res,* 35(Web Server issue), pp. W52-7.

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. and Taipale, J. (2018) 'CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response', *Nat Med,* 24(7), pp. 927-930.

Haeusler, A. R., Donnelly, C. J. and Rothstein, J. D. (2016) 'The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease', *Nat Rev Neurosci,* 17(6), pp. 383-95.

Harrington, L. B., Paez-Espino, D., Staahl, B. T., Chen, J. S., Ma, E., Kyrpides, N. C. and Doudna, J. A. (2017) 'A thermostable Cas9 with increased lifetime in human plasma', *Nat Commun,* 8(1), pp. 1424.

Hedlund, B. P. and Staley, J. T. (2002) 'Phylogeny of the genus Simonsiella and other members of the Neisseriaceae', *Int J Syst Evol Microbiol,* 52(Pt 4), pp. 1377-82.

Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D. and Marraffini, L. A. (2015) 'Cas9 specifies functional viral targets during CRISPR-Cas adaptation', *Nature,* 519(7542), pp. 199-202.

Hendel, A., Bak, R. O., Clark, J. T., Kennedy, A. B., Ryan, D. E., Roy, S., Steinfeld, I., Lunstad, B. D., Kaiser, R. J., Wilkens, A. B., Bacchetta, R., Tsalenko, A., Dellinger, D., Bruhn, L. and Porteus, M. H. (2015) 'Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells', *Nat Biotechnol,* 33(9), pp. 985-989.

Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L. F., Sontheimer, E. J. and Thomson, J. A. (2013) 'Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis', *Proc Natl Acad Sci U S A,* 110(39), pp. 15644-9.

Ibraheim, R., Song, C.-Q., Mir, A., Amrani, N., Xue, W. and Sontheimer, E. J. (2018a) 'All-in-One Adeno-associated Virus Delivery and Genome Editing by Neisseria meningitidis Cas9 in vivo', *bioRxiv*, pp. https://doi.org/10.1101/295055.
Ibraheim, R., Song, C. Q., Mir, A., Amrani, N., Xue, W. and Sontheimer, E. J. (2018b) 'All-in-one adeno-associated virus delivery and genome editing by Neisseria meningitidis Cas9 in vivo', *Genome Biol,* 19(1), pp. 137.
Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) 'Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product', *J Bacteriol,* 169(12), pp. 5429-33.
Iyer, S., Suresh, S., Guo, D., Daman, K., Chen, J. C. J., Liu, P., Zieger, M., Luk, K., Roscoe, B. P., Mueller, C., King, O. D., Emerson, C. P., Jr. and Wolfe, S. A. (2019) 'Precise therapeutic gene correction by a simple nuclease-induced double-stranded break', *Nature,* 568(7753), pp. 561-565.
Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C. and Brouns, S. J. (2017) 'CRISPR-Cas: Adapting to change', *Science,* 356(6333).
Jansen, R., Embden, J. D., Gaastra, W. and Schouls, L. M. (2002) 'Identification of genes that are associated with DNA repeats in prokaryotes', *Mol Microbiol,* 43(6), pp. 1565-75.
Jasin, M. and Haber, J. E. (2016) 'The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair', *DNA Repair (Amst),* 44, pp. 6-16.
Jiang, F. and Doudna, J. A. (2017) 'CRISPR-Cas9 Structures and Mechanisms', *Annu Rev Biophys,* 46, pp. 505-529.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science,* 337(6096), pp. 816-21.
Keeler, A. M. and Flotte, T. R. (2019) 'Recombinant Adeno-Associated Virus Gene Therapy in Light of Luxturna (and Zolgensma and Glybera): Where Are We, and How Did We Get Here?', *Annu Rev Virol,* 6(1), pp. 601-621.
Khvorova, A. and Watts, J. K. (2017) 'The chemical evolution of oligonucleotide therapies of clinical utility', *Nat Biotechnol,* 35(3), pp. 238-248.
Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., Burrell, J. R. and Zoing, M. C. (2011) 'Amyotrophic lateral sclerosis', *Lancet,* 377(9769), pp. 942-55.
Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H. Y., Song, D. W., Lee, K. J., Jung, M. H., Kim, S., Kim, J. H., Kim, J. H. and Kim, J. S. (2017) 'In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni', *Nat Commun,* 8, pp. 14500.
Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z. and Joung, J. K. (2015) 'Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition', *Nat Biotechnol,* 33(12), pp. 1293-1298.
Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. and Liu, D. R. (2016) 'Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage', *Nature,* 533(7603), pp. 420-4.

Koo, T., Lu-Nguyen, N. B., Malerba, A., Kim, E., Kim, D., Cappellari, O., Cho, H. Y., Dickson, G., Popplewell, L. and Kim, J. S. (2018) 'Functional Rescue of Dystrophin Deficiency in Mice Caused by Frameshift Mutations Using Campylobacter jejuni Cas9', *Mol Ther,* 26(6), pp. 1529-1538.
Koonin, E. V. and Makarova, K. S. (2009) 'CRISPR-Cas: an adaptive immunity system in prokaryotes', *F1000 Biol Rep,* 1, pp. 95.
Koonin, E. V. and Makarova, K. S. (2019) 'Origins and evolution of CRISPR-Cas systems', *Philos Trans R Soc Lond B Biol Sci,* 374(1772), pp. 20180087.
Kosicki, M., Tomberg, K. and Bradley, A. (2018) 'Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements', *Nat Biotechnol,* 36(8), pp. 765-771.
Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S. and Qi, L. S. (2013) 'CRISPR interference (CRISPRi) for sequence-specific control of gene expression', *Nat Protoc,* 8(11), pp. 2180-96.
Lau, C. H. and Suh, Y. (2017) 'In vivo genome editing in animals using AAV-CRISPR system: applications to translational research of human disease', *F1000Res,* 6, pp. 2153.
Lee, C. M., Cradick, T. J. and Bao, G. (2016) 'The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells', *Mol Ther,* 24(3), pp. 645-54.
Lee, J., Mir, A., Edraki, A., Garcia, B., Amrani, N., Lou, H. E., Gainetdinov, I., Pawluk, A., Ibraheim, R., Gao, X. D., Liu, P., Davidson, A. R., Maxwell, K. L. and Sontheimer, E. J. (2018) 'Potent Cas9 Inhibition in Bacterial and Human Cells by AcrIIC4 and AcrIIC5 Anti-CRISPR Proteins', *MBio,* 9(6).
Lesnik, E. A., Sampath, R., Levene, H. B., Henderson, T. J., McNeil, J. A. and Ecker, D. J. (2001) 'Prediction of rho-independent transcriptional terminators in Escherichia coli', *Nucleic Acids Res,* 29(17), pp. 3583-94.
Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., Lv, J., Xie, X., Chen, Y., Li, Y., Sun, Y., Bai, Y., Songyang, Z., Ma, W., Zhou, C. and Huang, J. (2015) 'CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes', *Protein Cell,* 6(5), pp. 363-372.
Lino, C. A., Harper, J. C., Carney, J. P. and Timlin, J. A. (2018) 'Delivering CRISPR: a review of the challenges and approaches', *Drug Deliv,* 25(1), pp. 1234-1257.
Liu, C., Zhang, L., Liu, H. and Cheng, K. (2017) 'Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications', *J Control Release,* 266, pp. 17-26.
Liu, X. S., Wu, H., Krzisch, M., Wu, X., Graef, J., Muffat, J., Hnisz, D., Li, C. H., Yuan, B., Xu, C., Li, Y., Vershkov, D., Cacace, A., Young, R. A. and Jaenisch, R. (2018) 'Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene', *Cell,* 172(5), pp. 979-992 e6.
Ma, H., Marti-Gutierrez, N., Park, S. W., Wu, J., Lee, Y., Suzuki, K., Koski, A., Ji, D., Hayama, T., Ahmed, R., Darby, H., Van Dyken, C., Li, Y., Kang, E., Park, A. R., Kim, D., Kim, S. T., Gong, J., Gu, Y., Xu, X., Battaglia, D., Krieg, S. A., Lee, D. M., Wu, D. H., Wolf, D. P., Heitner, S. B., Belmonte, J. C. I., Amato, P., Kim, J. S., Kaul, S. and

Mitalipov, S. (2017) 'Correction of a pathogenic gene mutation in human embryos', *Nature,* 548(7668), pp. 413-419.
Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., van der Oost, J. and Koonin, E. V. (2011) 'Evolution and classification of the CRISPR-Cas systems', *Nat Rev Microbiol,* 9(6), pp. 467-77.
Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., Backofen, R. and Koonin, E. V. (2015) 'An updated evolutionary classification of CRISPR-Cas systems', *Nat Rev Microbiol,* 13(11), pp. 722-36.
Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G. M. (2013) 'RNA-guided human genome engineering via Cas9', *Science,* 339(6121), pp. 823-6.
Marraffini, L. A. and Sontheimer, E. J. (2008) 'CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA', *Science,* 322(5909), pp. 1843-5.
Martino, A. T., Suzuki, M., Markusic, D. M., Zolotukhin, I., Ryals, R. C., Moghimi, B., Ertl, H. C., Muruve, D. A., Lee, B. and Herzog, R. W. (2011) 'The genome of self-complementary adeno-associated viral vectors increases Toll-like receptor 9-dependent innate immune responses in the liver', *Blood,* 117(24), pp. 6459-68.
Mir, A., Alterman, J. F., Hassler, M. R., Debacker, A. J., Hudgens, E., Echeverria, D., Brodsky, M. H., Khvorova, A., Watts, J. K. and Sontheimer, E. J. (2018a) 'Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing', *Nat Commun,* 9(1), pp. 2641.
Mir, A., Edraki, A., Lee, J. and Sontheimer, E. J. (2018b) 'Type II-C CRISPR-Cas9 Biology, Mechanism, and Application', *ACS Chem Biol,* 13(2), pp. 357-365.
Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. and Soria, E. (2005) 'Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements', *J Mol Evol,* 60(2), pp. 174-82.
Morgan, G. J., Hatfull, G. F., Casjens, S. and Hendrix, R. W. (2002) 'Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus', *J Mol Biol,* 317(3), pp. 337-59.
Muller, M., Lee, C. M., Gasiunas, G., Davis, T. H., Cradick, T. J., Siksnys, V., Bao, G., Cathomen, T. and Mussolino, C. (2016) 'Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome', *Mol Ther,* 24(3), pp. 636-44.
Musunuru, K. (2013) 'Genome editing of human pluripotent stem cells to generate human cellular disease models', *Dis Model Mech,* 6(4), pp. 896-904.
Nelson, C. E., Hakim, C. H., Ousterout, D. G., Thakore, P. I., Moreb, E. A., Castellanos Rivera, R. M., Madhavan, S., Pan, X., Ran, F. A., Yan, W. X., Asokan, A., Zhang, F., Duan, D. and Gersbach, C. A. (2016) 'In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy', *Science,* 351(6271), pp. 403-7.

Nishiyama, J., Mikuni, T. and Yasuda, R. (2017) 'Virus-Mediated Genome Editing via Homology-Directed Repair in Mitotic and Postmitotic Cells in Mammalian Brain', *Neuron,* 96(4), pp. 755-768 e5.
Norskov-Lauritsen, N. (2014) 'Classification, identification, and clinical significance of Haemophilus and Aggregatibacter species with host specificity for humans', *Clin Microbiol Rev,* 27(2), pp. 214-40.
Orthwein, A., Noordermeer, S. M., Wilson, M. D., Landry, S., Enchev, R. I., Sherker, A., Munro, M., Pinder, J., Salsman, J., Dellaire, G., Xia, B., Peter, M. and Durocher, D. (2015) 'A mechanism for the suppression of homologous recombination in G1 cells', *Nature,* 528(7582), pp. 422-6.
Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B. C., Barrangou, R. and Banfield, J. F. (2015) 'CRISPR immunity drives rapid phage genome evolution in Streptococcus thermophilus', *MBio,* 6(2).
Paulson, H. (2018) 'Repeat expansion diseases', *Handb Clin Neurol,* 147, pp. 105-123.
Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E. J., Maxwell, K. L. and Davidson, A. R. (2016) 'Naturally occurring off-switches for CRISPR-Cas9', *Cell,* 167(7), pp. 1829-1838 e9.
Pawluk, A., Bondy-Denomy, J., Cheung, V. H., Maxwell, K. L. and Davidson, A. R. (2014) 'A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa', *MBio,* 5(2), pp. e00896.
Pawluk, A., Davidson, A. R. and Maxwell, K. L. (2018) 'Anti-CRISPR: discovery, mechanism and function', *Nat Rev Microbiol,* 16(1), pp. 12-17.
Peters, O. M., Cabrera, G. T., Tran, H., Gendron, T. F., McKeon, J. E., Metterville, J., Weiss, A., Wightman, N., Salameh, J., Kim, J., Sun, H., Boylan, K. B., Dickson, D., Kennedy, Z., Lin, Z., Zhang, Y. J., Daughrity, L., Jung, C., Gao, F. B., Sapp, P. C., Horvitz, H. R., Bosco, D. A., Brown, S. P., de Jong, P., Petrucelli, L., Mueller, C. and Brown, R. H., Jr. (2015) 'Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice', *Neuron,* 88(5), pp. 902-909.
Petrov, D., Mansfield, C., Moussy, A. and Hermine, O. (2017) 'ALS Clinical Trials Review: 20 Years of Failure. Are We Any Closer to Registering a New Treatment?', *Front Aging Neurosci,* 9, pp. 68.
Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E. and Yuan, G. C. (2016) 'Analyzing CRISPR genome-editing experiments with CRISPResso', *Nat. Biotechnol.,* 34(7), pp. 695-7.
Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. and Lim, W. A. (2013) 'Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression', *Cell,* 152(5), pp. 1173-83.
Racanelli, V. and Rehermann, B. (2006) 'The liver as an immunological organ', *Hepatology,* 43(2 Suppl 1), pp. S54-62.
Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., Koonin, E. V., Sharp, P. A. and Zhang, F. (2015) 'In vivo genome editing using Staphylococcus aureus Cas9', *Nature,* 520(7546), pp. 186-91.

Rashid, S., Curtis, D. E., Garuti, R., Anderson, N. N., Bashmakov, Y., Ho, Y. K., Hammer, R. E., Moon, Y. A. and Horton, J. D. (2005) 'Decreased plasma cholesterol and hypersensitivity to statins in mice lacking Pcsk9', *Proc Natl Acad Sci U S A,* 102(15), pp. 5374-9.

Rauch, B. J., Silvis, M. R., Hultquist, J. F., Waters, C. S., McGregor, M. J., Krogan, N. J. and Bondy-Denomy, J. (2017) 'Inhibition of CRISPR-Cas9 with Bacteriophage Proteins', *Cell,* 168(1-2), pp. 150-158 e10.

Rees, H. A. and Liu, D. R. (2018) 'Base editing: precision chemistry on the genome and transcriptome of living cells', *Nat Rev Genet,* 19(12), pp. 770-788.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) 'The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli', *Nucleic Acids Res,* 39(21), pp. 9275-82.

Schumann, K., Lin, S., Boyer, E., Simeonov, D. R., Subramaniam, M., Gate, R. E., Haliburton, G. E., Ye, C. J., Bluestone, J. A., Doudna, J. A. and Marson, A. (2015) 'Generation of knock-in primary human T cells using Cas9 ribonucleoproteins', *Proc Natl Acad Sci U S A,* 112(33), pp. 10437-42.

Shi, Y., Lin, S., Staats, K. A., Li, Y., Chang, W. H., Hung, S. T., Hendricks, E., Linares, G. R., Wang, Y., Son, E. Y., Wen, X., Kisler, K., Wilkinson, B., Menendez, L., Sugawara, T., Woolwine, P., Huang, M., Cowan, M. J., Ge, B., Koutsodendris, N., Sandor, K. P., Komberg, J., Vangoor, V. R., Senthilkumar, K., Hennes, V., Seah, C., Nelson, A. R., Cheng, T. Y., Lee, S. J., August, P. R., Chen, J. A., Wisniewski, N., Hanson-Smith, V., Belgard, T. G., Zhang, A., Coba, M., Grunseich, C., Ward, M. E., van den Berg, L. H., Pasterkamp, R. J., Trotti, D., Zlokovic, B. V. and Ichida, J. K. (2018) 'Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons', *Nat Med,* 24(3), pp. 313-325.

Shin, J., Jiang, F., Liu, J. J., Bray, N. L., Rauch, B. J., Baik, S. H., Nogales, E., Bondy-Denomy, J., Corn, J. E. and Doudna, J. A. (2017) 'Disabling Cas9 by an anti-CRISPR DNA mimic', *Sci Adv,* 3(7), pp. e1701620.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O. O., Gootenberg, J. S., Makarova, K. S., Wolf, Y. I., Severinov, K., Zhang, F. and Koonin, E. V. (2017a) 'Diversity and evolution of class 2 CRISPR-Cas systems', *Nat Rev Microbiol,* 15(3), pp. 169-182.

Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V. and Koonin, E. V. (2017b) 'The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes', *MBio,* 8(5).

Stanley, S. Y. and Maxwell, K. L. (2018) 'Phage-Encoded Anti-CRISPR Defenses', *Annu Rev Genet,* 52, pp. 445-464.

Sternberg, S. H., LaFrance, B., Kaplan, M. and Doudna, J. A. (2015) 'Conformational control of DNA target cleavage by CRISPR-Cas9', *Nature,* 527(7576), pp. 110-3.

Sternberg, S. H., Richter, H., Charpentier, E. and Qimron, U. (2016) 'Adaptation in CRISPR-Cas Systems', *Mol Cell,* 61(6), pp. 797-808.

Swiech, L., Heidenreich, M., Banerjee, A., Habib, N., Li, Y., Trombetta, J., Sur, M. and Zhang, F. (2015) 'In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9', *Nat Biotechnol,* 33(1), pp. 102-6.

Taylor, J. P., Brown, R. H., Jr. and Cleveland, D. W. (2016) 'Decoding ALS: from genes to mechanism', *Nature,* 539(7628), pp. 197-206.
Thavalingam, A., Cheng, Z., Garcia, B., Huang, X., Shah, M., Sun, W., Wang, M., Harrington, L., Hwang, S., Hidalgo-Reyes, Y., Sontheimer, E. J., Doudna, J., Davidson, A. R., Moraes, T. F., Wang, Y. and Maxwell, K. L. (2019) 'Inhibition of CRISPR-Cas9 ribonucleoprotein complex assembly by anti-CRISPR AcrIIC2', *Nat Commun,* 10(1), pp. 2806.
Tsai, S. Q. and Joung, J. K. (2016) 'Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases', *Nat Rev Genet,* 17(5), pp. 300-12.
Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A. J., Le, L. P., Aryee, M. J. and Joung, J. K. (2014) 'GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases', *Nat. Biotechnol.,* 33, pp. 187-197.
Tycko, J., Myer, V. E. and Hsu, P. D. (2016) 'Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity', *Mol Cell,* 63(3), pp. 355-70.
van Houte, S., Ekroth, A. K., Broniewski, J. M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A. and Westra, E. R. (2016) 'The diversity-generating benefits of a prokaryotic adaptive immune system', *Nature,* 532(7599), pp. 385-8.
Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., Lander, E. S. and Sabatini, D. M. (2015) 'Identification and characterization of essential genes in the human genome', *Science,* 350(6264), pp. 1096-101.
Watters, K. E., Fellmann, C., Bai, H. B., Ren, S. M. and Doudna, J. A. (2018) 'Systematic discovery of natural CRISPR-Cas12a inhibitors', *Science,* 362(6411), pp. 236-239.
Wiese, J., Thiel, V., Gartner, A., Schmaljohann, R. and Imhoff, J. F. (2009) 'Kiloniella laminariae gen. nov., sp. nov., an alphaproteobacterium from the marine macroalga Laminaria saccharina', *Int J Syst Evol Microbiol,* 59(Pt 2), pp. 350-6.
Willems, A., Gilhaus, H., Beer, W., Mietke, H., Gelderblom, H. R., Burghardt, B., Voigt, W. and Reissbrodt, R. (2002) 'Brackiella oedipodis gen. nov., sp. nov., gram-negative, oxidase-positive rods that cause endocarditis of cotton-topped tamarin (Saguinus oedipus)', *Int J Syst Evol Microbiol,* 52(Pt 1), pp. 179-86.
Yamada, M., Watanabe, Y., Gootenberg, J. S., Hirano, H., Ran, F. A., Nakane, T., Ishitani, R., Zhang, F., Nishimasu, H. and Nureki, O. (2017) 'Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems', *Mol Cell,* 65(6), pp. 1109-1121 e3.
Yokoyama, T., Silversides, D. W., Waymire, K. G., Kwon, B. S., Takeuchi, T. and Overbeek, P. A. (1990) 'Conserved cysteine to serine mutation in tyrosinase is responsible for the classical albino mutation in laboratory mice', *Nucleic Acids Res,* 18(24), pp. 7293-8.
Yoon, Y., Wang, D., Tai, P. W. L., Riley, J., Gao, G. and Rivera-Perez, J. A. (2018) 'Streamlined ex vivo and in vivo genome editing in mouse embryos using recombinant adeno-associated viruses', *Nat Commun,* 9(1), pp. 412.

Zaiss, A. K., Liu, Q., Bowen, G. P., Wong, N. C., Bartlett, J. S. and Muruve, D. A. (2002) 'Differential activation of innate immune responses by adenovirus and adeno-associated virus vectors', *J Virol,* 76(9), pp. 4580-90.

Zarei, S., Carr, K., Reiley, L., Diaz, K., Guerra, O., Altamirano, P. F., Pagani, W., Lodin, D., Orozco, G. and Chinea, A. (2015) 'A comprehensive review of amyotrophic lateral sclerosis', *Surg Neurol Int,* 6, pp. 171.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V. and Zhang, F. (2015) 'Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system', *Cell,* 163(3), pp. 759-71.

Zhang, Y., Heidrich, N., Ampattu, B. J., Gunderson, C. W., Seifert, H. S., Schoen, C., Vogel, J. and Sontheimer, E. J. (2013) 'Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis', *Mol Cell,* 50(4), pp. 488-503.

Zhang, Y., Rajan, R., Seifert, H. S., Mondragón, A. and Sontheimer, E. J. (2015) 'DNase H activity of *Neisseria meningitidis* Cas9', *Mol. Cell,* 60, pp. 242-255.

Zhang, Z., Theurkauf, W. E., Weng, Z. and Zamore, P. D. (2012) 'Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection', *Silence,* 3(1), pp. 9.

Zhu, L. J., Holmes, B. R., Aronin, N. and Brodsky, M. H. (2014) 'CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems', *PLoS One,* 9(9), pp. e108424.

Zhu, L. J., Lawrence, M., Gupta, A., Pagés, H., Kucukural, A., Garber, M. and Wolfe, S. A. (2017) 'GUIDEseq: a bioconductor package to analyze GUIDE-Seq datasets for CRISPR-Cas nucleases', *BMC Genomics,* 18, pp. 379.

Zu, T., Liu, Y., Banez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T. M., Harms, M. B., Falchook, A. E., Subramony, S. H., Ostrow, L. W., Rothstein, J. D., Troncoso, J. C. and Ranum, L. P. (2013) 'RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia', *Proc Natl Acad Sci U S A,* 110(51), pp. E4968-77.

Zuris, J. A., Thompson, D. B., Shu, Y., Guilinger, J. P., Bessen, J. L., Hu, J. H., Maeder, M. L., Joung, J. K., Chen, Z. Y. and Liu, D. R. (2015) 'Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo', *Nat Biotechnol,* 33(1), pp. 73-80.